EQUITY RESEARCH | November 14, 2016

The following is a redacted version of the original report. See inside for details.

Goldman Sachs

Artificial intelligence is the apex technology of the information era. In the latest in our **Profiles in Innovation** series, we examine how advances in machine learning and deep learning have combined with more powerful computing and an ever-expanding pool of data to bring AI within reach for companies across industries. The development of AI-as-a-service has the potential to open new markets and disrupt the playing field in cloud computing. We believe the ability to leverage AI will become a defining attribute of competitive advantage for companies in coming years and will usher in a resurgence in productivity.

Heath P. Terry, CFA
(212) 357-1849
heath.terry@gs.com
Goldman, Sachs & Co.

Jesse Hulsing
(415) 249-7464
jesse.hulsing@gs.com
Goldman, Sachs & Co.

Mark Grant
(212) 357-4475
mark.grant@gs.com
Goldman, Sachs & Co.

Daniel Powell
(917) 343-4120
daniel.powell@gs.com
Goldman, Sachs & Co.

Piyush Mubayi
(852) 2978-1677
piyush.mubayi@gs.com
Goldman Sachs (Asia) L.L.C.

Waqar Syed
(212) 357-1804
waqar.syed@gs.com
Goldman, Sachs & Co.

AI

PROFILES IN INNOVATION

Artificial Intelligence

AI, Machine Learning and Data Fuel the Future of Productivity

The Goldman Sachs Group, Inc.

Contents

Contributing Authors: *Heath P. Terry, CFA, Jesse Hulsing, Robert D. Boroujerdi, Jan Hatzius, Piyush Mubayi, Mark Grant, Daniel Powell, Waqar Syed, Adam Hotchkiss, Komal Makkar, Yena Jeon, Toshiya Hari, Heather Bellini, CFA, Simona Jankowski, CFA, Matthew J. Fassler, Terence Flynn, PhD, Jerry Revich, CFA, Salveen Richter, CFA, Rob Joyce, Charles Long*

Note: The following is a redacted version of "Profiles in Innovation: Artificial Intelligence" originally published Nov. 14, 2016 [99pgs]. All company references in this note are for illustrative purposes only and should not be interpreted as investment recommendations.

Portfolio Manager's summary

Artificial Intelligence (AI) is the apex technology of the information age. The leap from computing built on the foundation of humans *telling* computers how to act, to computing built on the foundation of computers *learning* how to act has significant implications for every industry. While this moment in time may be viewed as the latest cycle of promise and disappointment before the next AI Winter (Exhibit 8), these investments and new technologies will at the very least leave us with the tangible economic benefit to productivity of machine learning.

In the meantime, AI, bots, and self-driving cars have risen to the forefront of popular culture and even political discourse. However, our research over the last year leads us to believe that this is not a false start, but an inflection point. As we shall explore in this report, the reasons for the inflection range from the obvious (more and faster compute and an explosion of more data) to the more nuanced (significant strides in deep learning, specialized hardware, and the rise of open source).

One of the more exciting aspects of the AI inflection is that "real-world" use cases abound. While deep-learning enabled advances in computer vision and such technologies as natural language processing are dramatically improving the quality of Apple's Siri, Amazon's Alexa, and Google's photo recognition, AI is not just "tech for tech". **Where large data sets are combined with powerful enough technology, value is being created and competitive advantage is being gained.**

For example, in healthcare, image recognition technology can improve the accuracy of cancer diagnosis. In agriculture, farmers and seed producers can utilize deep learning techniques to improve crop yields. In pharmaceuticals, deep learning is used to improve drug discovery. In energy, exploration effectiveness is being improved and equipment availability is being increased. In financial services, costs are being lowered and returns increased by opening up new data sets to faster analysis than previously possible. AI is in the very early stages of use case discovery, and as the necessary technology is democratized through cloud based services we believe a wave of innovation will follow, creating new winners and losers in every industry.

We interview GS Chief Economist Jan Hatzius about the impact AI/machine learning could have on lagging US productivity growth on p. 17.

The broad applicability of AI also leads us to the conclusion that it is a needle-moving technology for the global economy and a driver behind improving productivity and ending the period of stagnant productivity growth in the US. Leveraging the research of Chief GS economist Jan Hatzius, we frame the current stagnation in capital deepening and its associated impact on US productivity. We believe that AI technology driven improvements to productivity could, similar to the 1990s, drive corporates to invest in more capital and labor intensive projects, accelerating growth, improving profitability, and expanding equity valuations.

Implications

While we see artificial intelligence impacting every corporation, industry, and segment of the economy in time, there are four implications for investors that we see as among the most notable.

Productivity. AI and machine learning (ML) has the potential to set off a cycle of productivity growth that benefits economic growth, corporate profitability, returns on capital, and asset valuations. According to GS Chief Economist Jan Hatzius "In principle, it [AI] does seem like something that could be potentially captured better in the statistics than the last wave of innovation to the extent that artificial intelligence reduces costs and reduces the need for labor input into high value added types of production. Those cost saving innovations in the business sector are things statisticians are probably better set up to capture than increases in variety and availability of apps for the iPhone, for example. To the extent Artificial Intelligence has a broad based impact on cost structures in the business sector, I'm reasonably confident that it would be picked up by statisticians and would show up in the overall productivity numbers."

Premium technology. The value of speed in AI and machine learning has the potential to reverse the trend towards cheaper commodity hardware in building data centers and networks. We believe this could drive substantial shifts in market share in hardware, software, and services spending. For example, an AWS workload running on a "standard" datacenter compute instance costs as little as $0.0065/hour compared to $0.900 for a GPU instance optimized for AI.

Competitive Advantage. We see the potential for AI and machine learning to reshuffle the competitive order across every industry. Management teams that fail to invest in and leverage these technologies risk being passed by competitors that benefit from the strategic intelligence, productivity gains, and capital efficiencies they create.

New Company Creation. While we believe that much of the value in AI will accrue to large companies with the resources, data, and ability to invest, we expect that venture capitalists, entrepreneurs and technologists will continue to drive the creation of new companies that will, in turn, drive substantial innovation and value creation through, at the very least, M&A, though we certainly wouldn't dismiss the potential for a "Google or Facebook of AI" to emerge.

In the following pages we delve into AI the technology, its history, the ecosystem being created around machine learning, applications for these technologies across industries and the companies that are leading the way.

What is AI?

AI is the science and engineering of making intelligent machines and computer programs capable of learning and problem solving in ways that normally require human intelligence. Classically, these include natural language processing and translation, visual perception and pattern recognition, and decision making, but the number and complexity of applications is rapidly expanding.

In this report, we will focus most of our analysis on machine learning, a branch of AI, and deep learning, a branch of machine learning. We highlight two key points:

1. Simplistically, **machine learning** is algorithms that learn from examples and experience (i.e., data sets) rather than relying on hard-coded and predefined rules. In other words, rather than a developer telling a program how to distinguish between an apple and an orange, an algorithm is fed data ("trained") and learns on its own how to distinguish between an apple and an orange.

2. Major advances in **deep learning** are one of the driving forces behind the current AI inflection point. Deep learning is a sub-set of machine learning. In most traditional machine learning approaches, features (i.e., the inputs or attributes that may be predictive) are designed by humans. Feature engineering is a bottleneck and requires significant expertise. In unsupervised deep learning, **the important features are not predefined by humans, but learned and created by the algorithm.**

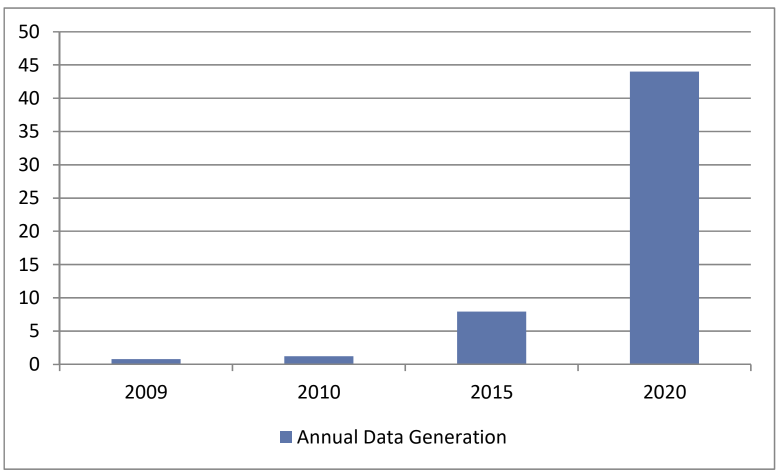
To be clear, we're not yet focusing on the kind of True, Strong, or General Artificial Intelligence that is meant to replicate independent human intelligence, and that is often the AI in popular culture. While there have been certain potential breakthroughs there, like Google DeepMind's AlphaGo system, which not only defeated a Go world champion, but did so using moves no human ever had before, we focus on the more immediately economically tangible areas of development in artificial intelligence.

Why is AI development accelerating now?

Major leaps in deep learning capabilities have been one of the catalysts behind the AI inflection currently underway. Neural networks, the underlying technology framework behind deep learning, have been around for decades, but three things have changed over the last five to ten years:

1. **Data.** There has been massive growth in the amount of unstructured data being created by the increasingly ubiquitous connected devices, machines, and systems globally. Neural networks become more effective the more data that they have, meaning that as the amount of data increases the number of problems that machine learning can solve using that data increases. Mobile, IoT, and maturation of low cost data storage and processing technologies (often in the cloud) has created massive growth in the number, size, and structure of available data sets. For example, Tesla has aggregated 780mn miles of driving data to date, and adding another million miles every ten hours through its connected cars, while Jasper (acquired by Cisco for $1.4bn in Feb. 2016) has a platform powering machine to machine communication for multiple automobile manufacturers and telco companies. Verizon made a similar investment in August when it announced it would acquire Fleetmatics, which connects remote sensors on vehicles to cloud software via increasingly fast wireless networks. The rollout of 5G will only accelerate the rate at which data can be generated and transmitted. Annual data generation is expected to reach 44 zettabytes (trillions of GB) by 2020, according to IDC's Digital Universe report, a CAGR of 141% over five years, suggesting that we are just beginning to see the use cases to which these technologies will be applied.

Exhibit 1: Annual data generation is expected to reach 44 zettabytes (44 trillion GB) by 2020, according to EMC/IDC
annual data generation globally (in ZB)



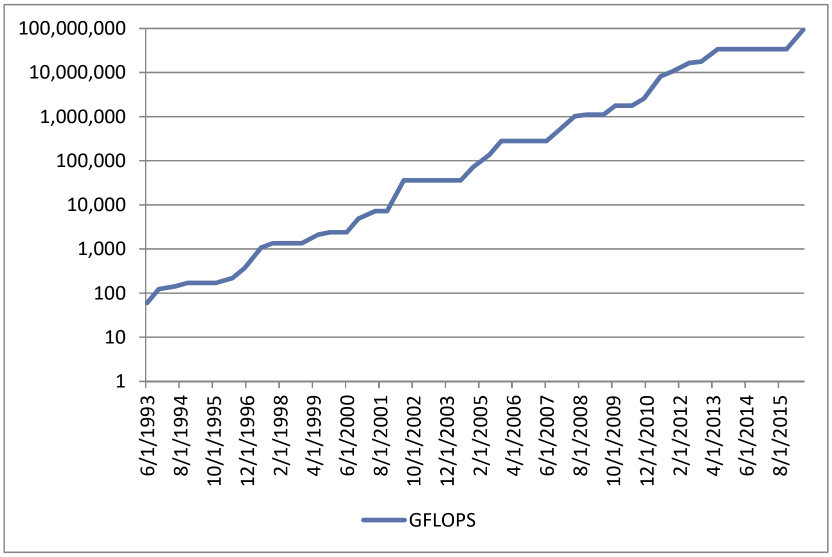
Source: EMC, IDC

2. **Faster hardware.** The repurposing of Graphic Processing Units (GPUs), the general availability of lower cost compute power, particularly through cloud services, and new models for building neural networks have dramatically increased the speed and accuracy of the results neural networks can produce. GPUs and their parallel architecture allow for faster training of machine learning systems compared to the traditional Central Processing Unit (CPU) based data center architecture. By repurposing graphics chips networks can iterate faster, leading to more accurate training in shorter periods of time. At the same time, the development of

specialized silicon, like the Field Programmable Gate Arrays being used by Microsoft and Baidu, allows for faster inference by trained deep learning systems. More broadly, the raw compute power of super computers has increased exponentially since 1993 (Exhibit 2). In 2016, a single high-end Nvidia video card for a gaming PC has sufficient compute power to have classified as the most powerful super computer in the world before 2002.

Exhibit 2: Raw compute performance of global supercomputers, measured in GFLOPs, has increased exponentially since 1993

Rpeak GFLOPS of #1 ranked global supercomputers on the Top500 list



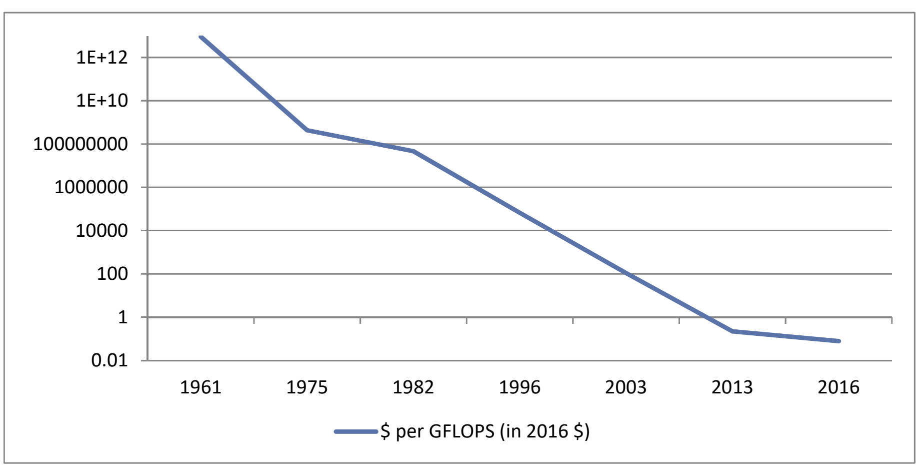
Source: top500.org, compiled by Goldman Sachs Global Investment Research

FLOPS (FLoating-point Operations Per Second): a measure of compute performance speed, wherein 1 GFLOPS equals 10^9 FLOPS and 1 TFLOPS equals 10^{12} FLOPS.

Costs of performance have also declined drastically. Nvidia's GPU (GTX 1080) delivers 9TFLOPS of performance for roughly $700, implying a price per GFLOPS of roughly 8 cents. In 1961, stringing together enough IBM 1620s to deliver a single GFLOPS of performance would require over $9 trillion (adjusted for inflation).

Exhibit 3: Price per unit of compute has decreased drastically over time

$ per GFLOPS in representative computer systems



Source: IBM, Cray, Sony, Nvidia, Press reports, Goldman Sachs Global Investment Research

3. **Better and more broadly available algorithms.** Better inputs (compute and data) are driving more R&D into algorithms to support deep learning use cases. Open source frameworks like Berkeley's Caffe, Google's TensorFlow, and Torch (used by Facebook) are allowing developers to compound their individual contributions by relying on tested base libraries as a foundation. As an example, TensorFlow, in less than a year, has become one of the most forked (or active) repositories on GitHub, the largest developer collaboration site. While not all AI is happening in a widely available, open source framework, (Apple is known for their secrecy in this field) their availability is certainly accelerating the development and open sourcing of more advanced tools.

Look around…

While the focus of this report is on where artificial intelligence is going and how companies are getting there, it is important to realize the extent to which AI is already impacting our lives.

Online Search. Just over a year ago, Google revealed that it had begun routing a significant number of its searches to RankBrain, an artificial intelligence system, making it one of the three most important signals, along with links and content, in Google's search algorithm.

Recommendation engines. Netflix, Amazon, and Pandora all use artificial intelligence to determine what movies to recommend, products to highlight, and songs to play. In May, Amazon open sourced, DSSTNE, the Deep Scalable Sparse Tensor Network Engine, "Destiny" for short, that it uses to produce product recommendations, so that it could be expanded beyond speech and language understanding and objection recognition.

Facial recognition. Both Google (FaceNet) and Facebook (DeepFace) have invested heavily in the technology necessary to identify with near 100 percent accuracy the faces in your photos. In January, Apple took a step further in buying Emotient, an AI startup that reads facial expressions to determine their emotional state. Clearly, these technologies are going far beyond tagging photos.

While there are countless additional consumer examples in personal assistants like Apple's Siri, credit and insurance risk scoring, and even weather prediction, in the coming pages we examine the way enterprises are using these technologies to accelerate growth, reduce costs, and control risk. At the rate these technologies and the applications for them are developing these will, at best, be a snapshot in time that provides some direction for the companies and investors working to stay in front of their competition.

What is Artificial Intelligence?

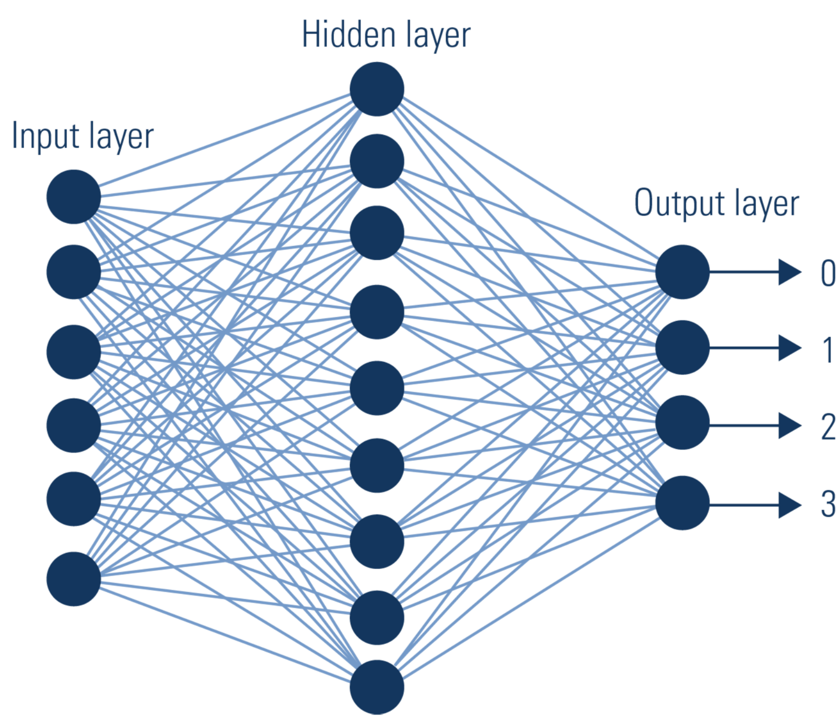
Artificial intelligence describes a science of simulating intelligent behavior in computers. It entails enabling computers to exhibit human-like behavioral traits including knowledge, reasoning, common sense, learning, and decision making.

What is machine learning? Machine learning is a branch of artificial intelligence and entails enabling computers to learn from data without being explicitly programmed. To provide simple context, a computer may be programmed to recognize trains in photos, but if it sees a photo of an object that only resembles a train (e.g. a museum built inside an old train, a toy train), a machine may falsely identify it as a train. In this scenario, machine learning would entail enabling the computer to learn from a large set of examples of trains and objects that only resemble trains, allowing it to better identify actual trains (thus achieving a level of artificial intelligence).

There are many real-world applications of machine learning. For instance, Netflix uses machine learning algorithms to generate personalized recommendations for users based on its massive volume of user behavior data and Zendesk uses customer interaction data to predict the likelihood of a customer being satisfied.

What is a neural network? A neural network in the context of AI/machine learning describes a type of computer architecture that simulates the structure of a human brain onto which AI/machine learning programs can be built. It consists of connected nodes in aggregate that can solve more complex problems and learn, like the neurons in a human brain.

Exhibit 4: Neural network
Multiple hidden layers would be characteristic of deep learning



Source: Michael A. Nielsen, "Neural Networks and Deep Learning", Determination Press, 2015, Goldman Sachs Global Investment Research

What is deep learning? Deep learning is a type of machine learning which entails training a hierarchy of "deep layers" of large neural networks, with each layer solving different aspects of a problem, allowing the system to solve more complex problems. Using the train example given above, deep learning system would contain layers that each identifies a different trait of a train. For example, the bottom layer would identify whether the object has windows. If the answer is yes, the next layer would look for wheels. The next would look for rectangular cars, so on and so forth, until the layers collectively identify the picture as a train or rejects it. Deep learning has been gaining popularity as a method of enhancing machine learning capabilities as technological advancements began to allow for training of large neural networks.

What is supervised learning? Unsupervised learning? Supervised and unsupervised learning are both types of machine learning. In supervised learning, the system is given a set of examples with "correct answers." Based on these examples, the system would learn to correctly predict the output based on what it has learned from the correct answers. Real-world applications of supervised learning include spam detection (e.g. the system may be fed a set of emails labeled "spam" and learn to correctly identify spam emails) and handwriting recognition. In unsupervised learning the system is not given correct answers, but unlabeled examples instead and left on its own to discover patterns. An example includes grouping customers into certain characteristics (e.g. purchasing frequency) based on patterns discovered from a large set of customer data.

What are some types of machine learning?

- **Classification.** Classify emails as spam, identify fraud, facial recognition, voice recognition, etc.
- **Clustering.** Comparing images, text or voice find similar items; identify clusters of unusual behavior.
- **Predictive.** Predict the likelihood of customer or employee churn based on web activity and other metadata; predict health issues based on wearable data.

What is General, Strong or True Artificial Intelligence? General, Strong, or True Artificial Intelligence are terms used for machine intelligence that fully replicates human intelligence including independent learning and decision making. While techniques like Whole Brain Emulation are being used to work towards the goal of General AI, the amount of compute power required is still considered far beyond current technologies, making General AI largely theoretical for the time being.

Key drivers of value creation

We believe profit pool creation (and destruction) related to the AI theme is best analyzed by first breaking AI down into four key inputs: **talent, data, infrastructure** and **silicon**. These inputs also double as barriers to adoption.

Talent

AI (deep learning in particular) is hard. Per our conversations with VCs and companies in the space, this has created a talent shortage and a competition for this talent among large internet and cloud computing vendors (Exhibit 5). AI talent is in high enough demand that "acquihires" are still a common means to acquire necessary talent. As the technology and tooling matures, talent may become less of a bottleneck. However, we believe talent will migrate to interesting, differentiated data sets**. Due to this, we believe large differentiated data sets are the most likely driver of growth and incremental profit dollars as we move into an AI-centric world.**

Exhibit 5: A Scarcity of AI Talent is Driving M&A

Target	Date Announced/ Reported	Description
Acquirer: AMZN		
2lemetry Inc.	3/16/2015	IoT: track and manage connected devices
Orbeus	Fall 2015	Photo-re cognition technology based on AI
Acquirer: AAPL		
Vocal IQ	10/2/2015	Speech recognition based on AI
Perceptio	10/5/2015	Uses AI to classify photos on smartphones
Emotient	1/7/2016	Uses AI to read people's emotions by analyzing facial expressions
Turi	8/5/2016	Machine learning platform for developers and data scientists
Tuplejump Software	9/26/2016	Machine learning/big data technology developer
Acquirer: CRM		
Tempo AI	5/29/2015	AI-based smart calendar app
MinHash	12/16/2015	Developed an AI platform and a virtual personal assistant (AILA) for marketers
PredictionIO	2/19/2016	Developed an open source-based machine learning server
Metamind	4/4/2016	Deep Learning platform (natural language processing, computer vision, database predictions, etc)
Acquirer: MSFT		
Equivio	1/21/2015	Machine-learning based text analytics service for legal and compliance
Revolution Analytics	1/27/2015	Open-source analytics company that specizlies in R programming language for statistical computing
Wand Labs	6/16/2016	Messaging app/chat bot devleoper
Genee	8/23/2016	AI scheduling assistance services
Acquirer: GOOGL		
DeepMind	1/16/2014	AI firm that specializes in machine learning, advanced algorithms, systems neuroscience
Emu	8/6/2014	Mobile messaging app with an AI assistant
Jetpac	8/15/2014	AI-based mobile photo app
Dark Blue Labs	10/23/2014	Deep learning startup specializing in understanding natural language
Vision Factory	10/23/2014	Deep learning startup specializing in visual recognition systems
Timeful	5/4/2015	Machine-learning based scheduling tool
Speaktoit (Api.ai)	9/21/2016	Speech recognition and natural language understanding solutions

Source: Bloomberg, company data, FactSet, The Guardian, Techcrunch, VentureBeat

Data

Data is *the* key input for AI. Deep learning effectiveness in particular is linked to larger data sets, as larger data sets prevent models from becoming over-fitted. For example, researchers from the Department of Radiology at Massachusetts General Hospital and Harvard Medical School used a convolutional neural network to identify CT images, assessing accuracy of the neural network based on training data size. As the training size grew larger, accuracy improved (Exhibit 6).

Exhibit 6: Medical Imaging (Body Part Image Recognition)
Training Size is Correlated With Accuracy; 0= least accurate, 100= most accurate

Training Data Size	5	10	20	50	100	200
Brain	0.3	3.39	45.71	59.07	72.82	98.44
Neck	21.3	30.63	79.97	99.34	99.74	99.33
Shoulder	2.98	21.39	69.64	86.57	95.53	92.94
Chest	23.39	34.45	62.53	96.18	95.25	99.61
Abdomen	0.1	3.23	35.4	65.83	91.01	95.18
Pelvis	0	1.15	15.99	55.9	83.7	88.45
Average	8.01	17.37	51.54	77.15	89.68	95.67

Source: Department of Radiology at Massachusetts General Hospital and Harvard Medical School

Most deep learning today is either supervised or semi-supervised, meaning all or some of the data utilized to train the model must be labeled by a human. Unsupervised machine learning is the current "holy grail" in AI, as raw un-labeled data could be utilized to train models. Broad adoption of deep learning will likely be tied to growth in large data sets (which is happening due to mobile and IoT) and to advances in unsupervised machine learning. However, we believe large differentiated data sets (electronic health records, omics data, geological data, weather data, etc.) will likely be a core driver of profit pool creation over the next decade.

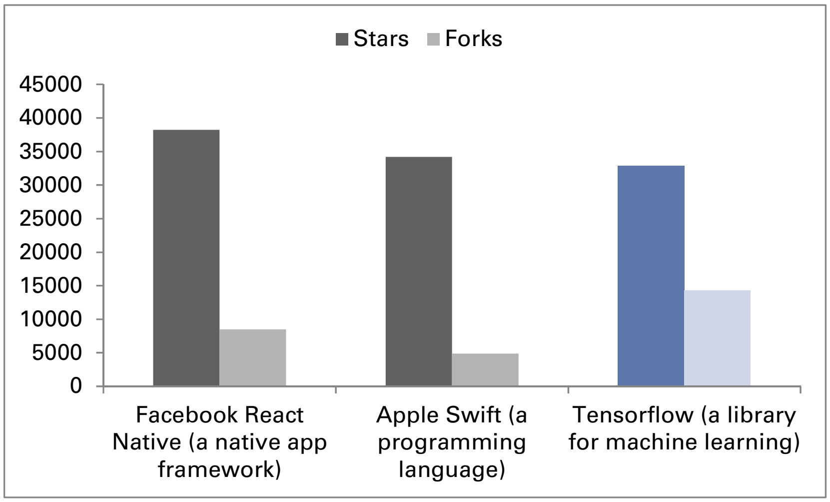
The amount of information created worldwide is expected to increase at a CAGR of 36% through 2020, reaching 44 Zettabytes (44 billion GB), according to IDC. Increases in connected devices (consumer and industrial), machine-to-machine communication, and remote sensors are combining to create large data sets that can then be mined for insights and to train adaptive algorithms. Availability of data has also increased dramatically in the last decade, with census, labor, weather, and even genome data available for free online in large quantities.

We are also seeing increased availability of satellite imagery, which requires a great deal of compute to fully analyze. The US Geological Survey's Landsat 7 and Landsat 8 satellites image the entire Earth every 8 days, and the USGS makes those images available for free – though even when compressed, the ultra-high definition images are approximately 1GB each, in file size. Other companies, like Orbital Insights are aggregating image data and creating commercial solutions across multiple industries.

Infrastructure

Hardware and infrastructure software are necessary to make AI work. We believe infrastructure to support AI will rapidly become commoditized. This view is based on two observations: 1) cloud computing vendors are well positioned to extend their offerings into AI infrastructure, 2) open source (TensorFlow, Caffe, Spark, etc.) has emerged as the primary driver of software innovation in AI. To spur adoption of AI, we believe large cloud vendors will continue to open source infrastructure capabilities, limiting the potential for profit pool creation.

Exhibit 7: Internet Giants (such as Google) are spurring interest in AI via open sourcing technologies (such as TensorFlow)
GitHub repositories most starred 2015-2016



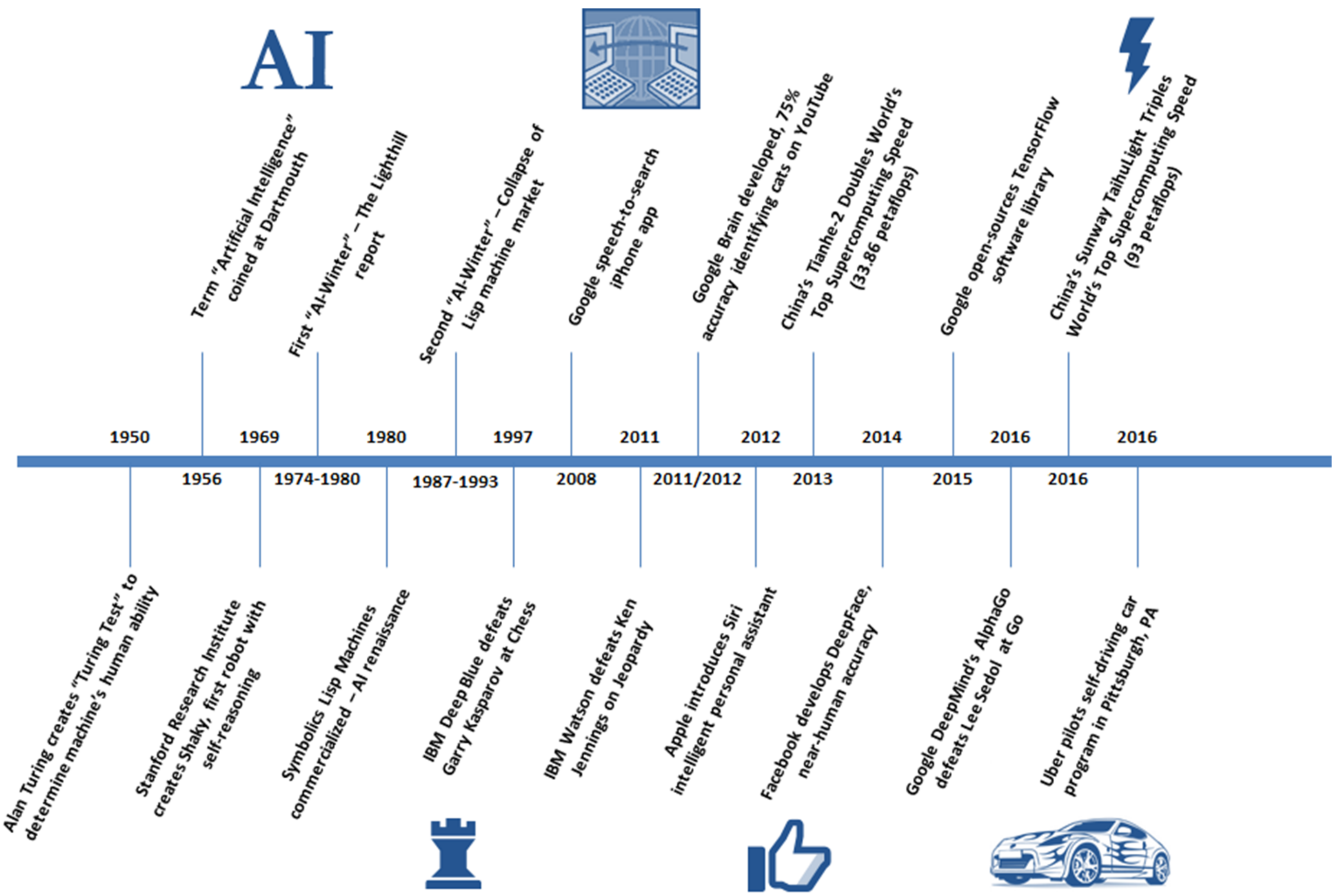
Source: GitHub

Silicon

The repurposing of GPUs for deep learning has been one of the key drivers of our current "AI Spring". Within the AI/ML ecosystem, there are two primary applications that determine the performance of a neural network with each requiring a different resource setup. The first is the construction and use of a training algorithm. The training algorithm leverages a large (usually the larger, the better) data set to find correlations and build a model that can determine the probability of an output, given a new input. Training is very resource-intensive, and most modern training is done on GPU-powered systems.

The use of models and algorithms once they have been trained is referred to as inference. Inference requires far less computing power, and typically combs through smaller, incremental data input sets. While some GPUs are optimized for inference (Nvidia's P4 series and M4 series, for example) given the single-purpose nature of inference, specialized silicon is being developed specifically for that application, referred to as FPGAs (Field Programmable Gate Array) and ASICs (Application Specific Integrated Circuit). This type of integrated circuit was originally developed for prototyping CPUs, but is increasingly being used for inference in artificial intelligence. Google's Tensor Processing Unit, is an example of an ASIC purpose-built for AI and machine learning. Microsoft has been using FPGA chips for inference, as well. Intel acquired FPGA manufacturer, Altera, in 2015 on the view that by 2020, a third of data centers could be leveraging FPGAs for specialized use cases. Xilinx, which pioneered commercially viable FPGAs in the 1980s, has pointed to the cloud and large data centers as a significant avenue of growth going forward, having announced a strategic customer relationship with Baidu. Data centers make up roughly 5% of Xilinx's revenue now.

Exhibit 8: Evolution of AI: 1950-Present



Source: Company data, Goldman Sachs Global Investment Research

Fueling the future of productivity

Labor productivity growth in the U.S. has come to a halt in recent years after modest growth in the past decade and significant growth in the mid-late 1990s. We believe that proliferation of consumable machine learning and AI has the potential to dramatically shift the productivity paradigm across global industries, in a way similar to the broad scale adoption of internet technologies in the 1990s.

Across industries, we see a roughly 0.5%-1.5% reduction in labor hours spurred by automation and efficiency gains brought to bear by AI/ML technologies resulting in a +51-154bps impact on productivity growth by 2025. While we expect AI/ML to improve both the denominator and numerator of productivity over time, we believe the most significant, early impacts will be on the automation of lower-wage tasks – driving similar levels of output growth with less labor hours. Our base case AI/ML driven improvement of 97 bps implies a 2025 productivity growth IT contribution of 1.61%, or 11bps higher than 1995-2004 (Exhibits 9, 10).

Exhibit 9: Productivity analysis

$ millions, assumes linear nominal GDP growth beyond 2019

US	2016E	2017E	2018E	2019E	2020E	2021E	2022E	2023E	2024E	2025E
Output										
US Nominal GDP* ($bn)	18,552	19,300	20,045	20,757	21,470	22,183	22,895	23,608	24,321	25,034
	2.9%	4.0%	3.9%	3.6%	3.4%	3.3%	3.2%	3.1%	3.0%	2.9%
Productivity										
Labor productivity	69.0	70.4	71.8	73.1	74.3	75.4	76.5	77.6	78.6	79.7
yoy growth (%)	0.9%	2.1%	2.0%	1.7%	1.6%	1.6%	1.5%	1.4%	1.3%	1.3%
Labor hours (mn)	268,958	273,992	279,026	284,060	289,094	294,128	299,162	304,196	309,230	314,264

ML/AI impact	Low	Base	High
Labor hours reduction (mn)	(1,571)	(2,969)	(4,714)
Reduction	-0.5%	-1%	-1.5%
2025 Labor hours (mn)	312,693	311,295	309,550
2025 GDP ($bn)	25,034	25,034	25,034
Labor productivity	80.1	80.4	80.9
yoy growth (%)	*1.8%*	*2.2%*	*2.8%*
Improvement (bps)	***51***	***97***	***154***

Source: OECD, US Bureau of Labor Statistics, Goldman Sachs Global Investment Research

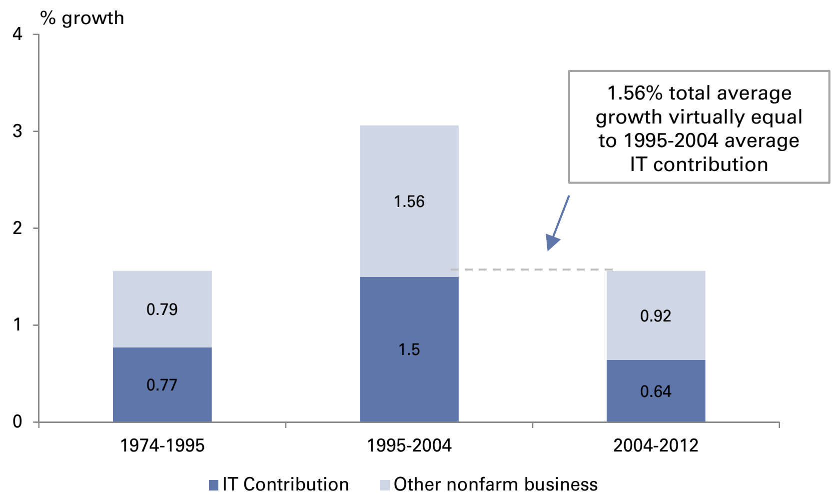
Technology and productivity growth

The 1990s technology boom saw unusual amplification of each of the two primary components of productivity, capital deepening and multifactor productivity (MFP), and was strongly correlated with rising equity valuations.

Capital Deepening. GS economist Jan Hatzius has provided recent analysis on the anti-cyclical tendency of capital deepening (capital stock per labor hour), as labor hours historically tend to rise during expansionary periods without an equal surge in capital stock (see Jan's report: *"Productivity Paradox v2.0 Revisited"*, published on 09/2/2016). In the 1990's, capital deepening increased markedly, highlighted by atypical capital investment increases that outpaced growth in the labor market.

Multifactor productivity (MFP). A March, 2013 Federal Reserve study by David Byrne et al. suggests that the simultaneous diffusion of technology into IT-producing and general operations processes contributed to creating a threefold spike in growth (output per labor hour) during the 1990s, with IT-producing sectors responsible for at most 49% of the average annual increase in annual productivity growth from the pre-boom period to the period between 1995 and 2004 (Exhibit 10).

Exhibit 10: Late 90s: IT-producing sectors contribute nearly half of productivity growth
But lose value and share in growth post-tech boom



Source: Federal Reserve Board, Goldman Sachs Global Investment Research

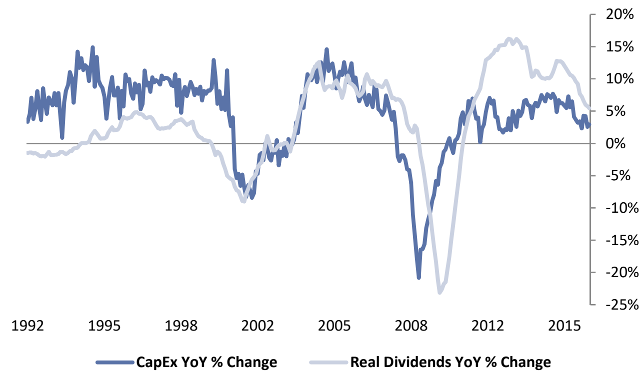
Post-millennium stagnation. During the past decade, capital deepening growth related to IT applications (computer hardware, software, and telecom) has stagnated. IT capital, relative to broader market capital, has contributed less to overall growth in this component than average contributions during and even before the tech boom. Aggregate labor hours have been increasing, but the contribution of capital intensity to productivity has drastically underperformed versus the 1990s. The introduction of increasingly sophisticated, consumable machine learning and AI may be a catalyst in bringing capital intensity back to the forefront, in our view, significantly increasing the productivity of labor similar to the cycle we saw in the 1990's.

We're more optimistic on the MFP side of the equation. GS economists have highlighted (*Productivity Paradox v2.0 Revisited*, 9/2/2016) that upward biases on ICT prices and a growth in inputs towards unmonetized outputs (free online content, back-end processes, etc.) add to the understatement of real GDP and productivity growth. Evolution of internet giants like Facebook and Google highlight the idea that complex input labor and capital aren't necessarily converted into traditional consumer product monetization captured in standard productivity metrics.

AI/ML induced productivity could impact investment

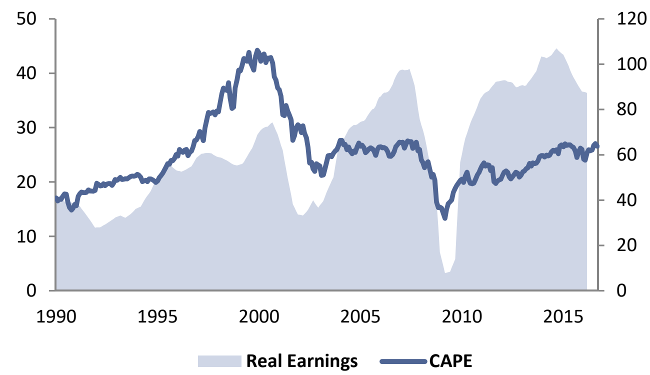
We believe that one of the potential impacts of increasing productivity from AI/ML could be a shift in the way companies allocate capital. Since mid-2011, the growth in dividends and share repurchases has significantly exceeded capex growth, as reluctance among management teams to investment in capital projects remains post-recession.

Exhibit 11: Companies are hesitant to sacrifice dividends
Clear shift in cash utilization strategy



Source: Shiller S&P Analysis, U.S. Department of Commerce, Goldman Sachs Global Investment Research

Exhibit 12: Cyclically adj. P/E ratios in a sluggish recovery
Valuations only just hitting pre-recession levels



Source: Shiller S&P Analysis, Goldman Sachs Global Investment Research

Productivity increases have the potential to renew management confidence and encourage companies to invest in productive capital similar to the 1990s, where yoy capex growth, measured by our GS capex tracker, consistently outpaced yoy dividend growth as measured in Yale Professor Robert Shiller's S&P 500 analysis (Exhibit 11). We further believe that investors would value such a shift with the support of productivity gains. Cyclically adjusted P/E ratios underwent significant inflation during this period of capex investment and related productivity growth, while current valuations have only just reached pre-recession levels (Exhibit 12).

AI & The Productivity Paradox: An interview with Jan Hatzius

Jan Hatzius

Heath Terry

GS Research Internet analyst Heath Terry sat down with Chief Economist Jan Hatzius to discuss the role AI & machine learning could play in boosting lagging labor productivity.

Heath Terry: What has led to the lack of measurable productivity growth over the last decade?

Jan Hatzius: A good starting point is the 1990s, where we did have a sizeable measured productivity acceleration which was mainly technology driven. The technology sector had gotten bigger and measured output growth in the technology sector was very rapid, and that was enough to produce an overall acceleration in the economy. Recently, over the last 10 years or so, we've seen a renewed deceleration to productivity growth rates that are as low as the 1970s and 1980s, and potentially even lower.

I think there is more than one driver, but I would say there are three things that have lowered productivity growth in my view. One is a bit of a cyclical effect. We still have some hangover from the great recession with a relatively slow pace of capital accumulation, relatively low levels of investment, and relatively rapid employment growth. Since productivity is measured as output per hour worked, that somewhat perversely means that you can have lower productivity numbers when the labor market is improving rapidly.

Another factor may be some slowdown in the overall pace of technological progress. It's reasonable to believe that perhaps the 1990s, with the introduction of the internet, was a relatively rapid period for technological progress, and I would say there is some support for the idea that it is a little slower now.

The third point is that technological progress that has occurred over the last decade, like mobile and consumer focused technology, is something that statisticians are not very well set up to capture in the numbers. The reason is that quality improvement in a lot of the new technologies that have dominated in the last decade or so is very difficult to capture in a quantitative sense. Statisticians are not building in significant quality improvement into a lot of the areas that have been at the cutting edge.

So I would point to three things. There are cyclical effects, probably some slowdown in technological progress, and very likely an increase in the measurement error in the productivity statistics.

Terry: Back to the productivity boom in the 1990s, what role did technology play?

Hatzius: What drove it was mainly general purpose technologies like semiconductors and computers, which had become much larger as a share of the economy than they were in the 1970s or 1980s and where technological progress was very rapid, in ways that statisticians were well set up to measure. The statisticians in the 1990s had made a lot of effort to capture quality improvement faster; processer speeds, more memory, better attributes in computer hardware, which led to large increases in the measured contribution of the technology sector. The technology sector was very central to pick up in the productivity numbers from the 1990s lasting to the early and mid-2000s.

Terry: We've seen a lot of technology development over the last 10-15 years. Why hasn't there been a similar impact to productivity from technologies such as the iPhone, Facebook, and the development of cloud computing?

Hatzius: We don't have a full answer to it, but I do think an important part of the answer is the statistical ability to measure improvement in quality, and the impact of new products in the economic statistics is limited. It's relatively easy to measure nominal GDP, that's basically a matter of adding up receipts. There is room for measurement error as there is in almost everything, but I don't have real first order concern that measurement is getting worse in terms of measuring nominal GDP. Converting nominal GDP numbers into real GDP numbers by deflating it with a quality adjusted overall price index is where I think things get very difficult. If you look, for example, at the way software sectors enter the official numbers, if you believe the official numbers, $1000 of outlay on software now buys you just as much real software as $1000 of outlay bought you in the 1990s. There has been no improvement in what you get for your money in the software sector. That's just very difficult to believe. That just does not pass the smell test. Because of the difficulty of capturing these quality improvements, the fact that the technology sector has increasingly moved from general purpose hardware to specialized hardware, software and digital products has led to increased understatement and mismeasurement.

Terry: What kind of impact could the development of technologies like artificial intelligence and machine learning have on productivity?

Hatzius: In principle, it does seem like something that could be potentially captured better in the statistics than the last wave of innovation to the extent that artificial intelligence reduces costs and reduces the need for labor input into high value added types of production. Those cost saving innovations in the business sector are things statisticians are probably better set up to capture than increases in variety and availability of apps for the iPhone, for example. To the extent Artificial Intelligence has a broad based impact on cost structures in the business sector, I'm reasonably confident that it would be picked up by statisticians and would show up in the overall productivity numbers.

I would just add one general caveat, which is that the U.S. economy is very large. Even things that are important in one sector of the economy and seem like an overwhelming force often look less important when you divide by $18tn, which is the level of U.S. nominal GDP, so the contribution in percentage terms may not look as large as one might think from a bottom up perspective. But in principle, this is something that could have a measurable impact.

Terry: You touched on the impact to cost, how do you see something like that impacting pricing? Does that become a contributor to the broader deflationary force that we've seen in certain parts of the economy?

Hatzius: I certainly think that in terms of productivity gains, the first order impact is often to lower costs and lower prices. Keeping everything else constant would mean lower overall inflation in the economy. It's often not the right assumption that you want to keep everything else constant; there are economic policy makers and there is a Federal Reserve, and if the impact is large then the Federal Reserve is going to run an easier monetary policy and allow people who may no longer be working in areas affected by artificial intelligence to find jobs elsewhere. There may be non-inflationary running room for the Fed to do that. In the longer term, we generally don't find that cost saving innovations lead to higher unemployment rates or significantly lower inflation. In the short-run that may be the consequence, but in the longer term when policy reacts, the economy ends up at similar unemployment and inflation levels.

Terry: Those themes that emerge out of this: AI taking jobs or robotics taking over labor, is that something over time you don't seem as being legitimate?

Hatzius: These fears have been around for many years, and what I think we can say is that so far they haven't really been borne out. If we go back to the early 19th century, there were worries about mechanized spinning machines and the idea that this would put large numbers of people out of work. In the short run, that disruption is something that can have a significant impact, but it's not the case that technological progress over the longer stretch of history has led to higher unemployment rates. That is not the case. My best guess is not that we would end up with much higher unemployment rates, because in the end people are going to find something that needs to be done that requires humans and human labor. My expectation is that it could be tumultuous impact but I don't think it is something that will leave us with a higher unemployment rate.

Terry: Over the past decade, we've seen corporate profits increasingly going to buybacks and dividends over capital investment. Is there a threshold where from a macro-economic perspective productivity needs to be in order to drive investment and capital?

Hatzius: Investment and productivity are linked, and causality goes in both directions. In recent years, we've had relatively low levels of investment largely for cyclical reasons because there was still a lot of spare capacity in the economy and capital stock was underutilized after the great recession. There wasn't a strong economic incentive to invest in new capacity. I think that is on the mend, we have seen some pickup in investment rates. There is a bigger contribution from business investment to productivity growth in the next couple of years than there was in, say, 2010 and 2011. In terms of the causality in the other direction, the opportunity for productivity growth is a big driver of investment, depending on the sort of discoveries that are being made in cutting edge sectors. It seems like there are still some significant discoveries and if that continues to be the case, then there will also be an incentive to invest.

Terry: When we see gains in productivity historically, how do those typically impact corporate profits? Do costs simply move to another part of the income statement as companies seek competitive advantage or do we actually see sustainable increases in profitability?

Hatzius: My reading of the historical evidence is that initially a productivity improvement falls to the bottom line in the company that has that opportunity, but eventually those high returns get competed away because more entrants try to get a piece of the action. It can be sustainable for some period of time, but over the longer term presumably if the market mechanism is working it will be competed away.

Terry: To the extent that we see technology driven improvements and efficiency, what impact do you tend to think that has on asset valuation? In the 90s, we saw a related market reaction to the productivity that you were talking about, what is the potential for something like that to repeat itself to the extent that we see this kind of productivity improvement around Artificial Intelligence and machine learning?

Hatzius: As far as the overall economy is concerned, I do think that if you had evidence of a more sustained productivity reacceleration and if you found that a lot of the fears that surfaced in recent years that were stuck with this subpar productivity growth pace went away, I think you would probably see a revaluation of equities.

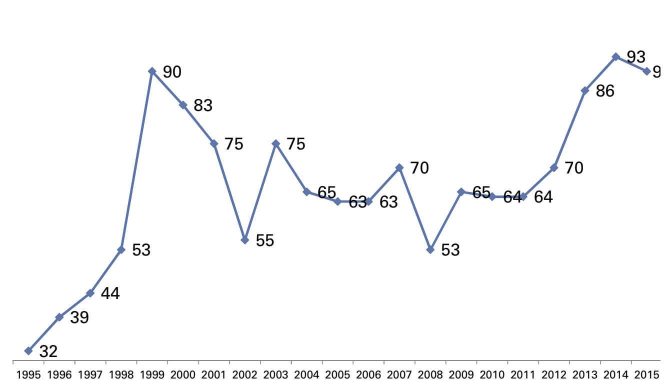
In particular, keeping all else the same, we find that periods of faster productivity growth also mean periods of higher asset valuations. If we look at the 1990s, we did see that. We did have a large bubble that developed towards the end of that period and the aftermath was quite painful. These things can be temporary, but I think we do typically see a revaluation.

The Ecosystem: Cloud services, open source key beneficiaries of the coming investment cycle in AI

We believe the ability to leverage AI technologies will become one of the major defining attributes of competitive advantage across all major industries in the coming years. While the strategy will differ by company size and industry, management teams that don't focus on leading in AI and benefiting from the resulting product innovation, labor efficiencies, and capital leverage risk being left behind. Accordingly, we believe the need for companies to invest in these new technologies to stay competitive will drive a boom in demand for the talent, services, and hardware underlying artificial intelligence.

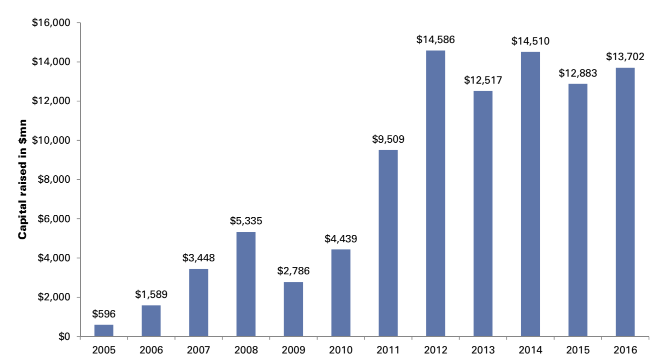
As a comparison, the 1990s tech-driven productivity boom drove a corresponding boom in enablers. Increased capital spending on technology drove an increase in business formation to capture this capital spending. Software, hardware and networking company formation inflected, before the inevitable industry consolidation occurred. Exhibit 13 below highlights this pattern within the software industry. The number of public software companies with between $200mn and $5bn in inflation adjusted market capitalization nearly tripled in the 1995-1999 period, before consolidating in the mid-2000s.

Exhibit 13: Rapid growth in the enabler ecosystem accompanied the 1990s productivity boom
\# of software companies with inflation adjusted market cap $200mn-$5bn (2015 dollars)



Source: Factset, Goldman Sachs Investment Research

Exhibit 14: Venture capital investing in AI has inflected this decade
Total VC funding in AI/ML by year

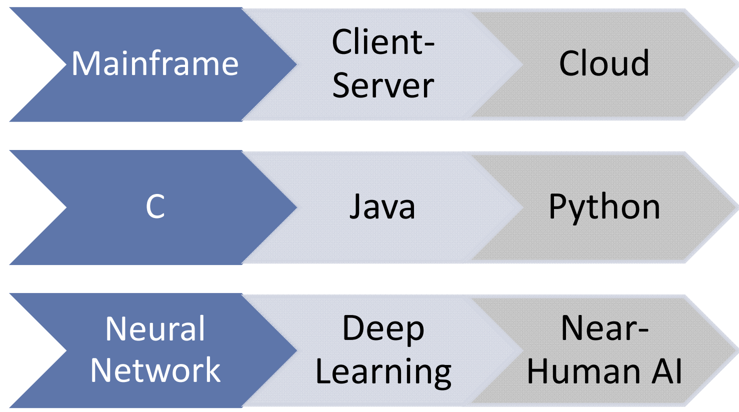


Source: PitchBook

We see potential for a similar boom related to the coming cycle of AI driven productivity, with value being created across software, hardware, data, and services providers as enterprises invest to capitalize on the potential of AI. Reflecting this opportunity, and as highlighted in Exhibit 14 above, VC funding into AI-related start-ups has inflected sharply this decade. The potential for a coming boom in enterprise AI investment has also started to drive consolidation. Cloud platforms in particular have invested heavily in AI talent., with Google, Amazon, Microsoft and Salesforce making a combined 17 AI-related acquisitions since 2014 (Exhibit 5).

We also see some benefit in the contextual comparison of where we are in the development of AI and ML technologies to historical technology cycles. As with other major technology cycles of the last 50 years, compute (and Moore's Law) has been both the inhibitor and the enabler of progress. For example, in systems architecture we have witnessed an evolution that began with the mainframe system, which then ceded to the client-server model, and has begun to be displaced by a cloud/mobile paradigm in recent years. The driver of this evolution was improvements in compute, storage capacity, and bandwidth. Each transition came with an accompanying shift in how applications were developed including the advent and evolution of various new programming languages (see Exhibit 15) and the type of applications that were possible. In context, AI has existed as a concept for decades, with conceptual neural networks emerging in the 1960s, though compute power was insufficient to allow for any practical use cases until recent years. We believe we are in early days of the AI platform, analogous to the initial commercialization of the mainframe in the 1950s and the commercialization of the smartphone and cloud in the 2000s. As the platform curve inflects (which we think is happening) an explosion of apps, tools, and services enablers emerge, which we discuss in more detail below.

Exhibit 15: AI advances can be compared to historical technological evolutions in systems architecture and programming language adoption, though we believe we are still in very early stages of development and adoption



Source: Goldman Sachs Global Investment Research

Enablers are evolving along three planes: DIY, Services and AI-aaS

As outlined in the following sections, we are beginning to see leaders and investment emerge along three planes:

1. **Do-it-yourself enablement** – Enterprises with talent and differentiated data will likely invest heavily in machine learning capabilities. To support these efforts, we are seeing a new "AI stack" emerging. The AI stack has similar components to the historical compute stack: silicon, storage, infrastructure software, data processing engines, programming languages and tools. As we will walk through below, the inputs to the AI stack are mostly a combination of open source (from providers such as Databricks, Cloudera, Hortonworks and Skymind) and services provided by cloud platforms such as Microsoft, Google, Amazon and Baidu.

2. **Consulting services** – Many organizations will have unique data sets and a desire to build AI services for internal usage, customers, and partners. Because AI talent is currently a scarce resource, professional services providers are emerging to help bridge the gap. Newer models are also emerging. Kaggle, as an example, connects organizations with thousands of data scientists to help solve AI-related problems.

3. **AI-as-a-service (AI-aaS)** – We see the potential for a significant amount of innovation and new market creation in a category we call AI-aaS. AI-aaS is likely to

develop on multiple fronts, but the core idea is that rather than training their own deep learning systems, many enterprises will instead access trained deep learning systems from outside providers. An example of AI-aaS would be image APIs from start-up Clarifai and from Google. A developer using image recognition in an application would call the Vision API each time image recognition is required in an application. Similar AI-aaS offerings are likely to be developed by SaaS providers who have unique horizontal data sets, startups addressing niches where data and talent are scarce (medical imaging being an example), and companies that have differentiated data that might be valuable to suppliers, customers or partners.

DIY: Cloud Platforms and Open Source Likely to be The Picks and Shovels of AI

Machine learning (and deep learning in particular) remains firmly in the Innovator/Early Adopter segments of the market compared to the rapid advances in AI. Based on discussions with companies and VCs in the space, we believe AI/ML is being used heavily by internet companies, industry-focused services providers (such as the Broad Institute) and a tail of larger Fortune 500 organizations (with emerging use cases highlighted in our industry vignettes).

The biggest barriers to adoption today are data and talent. However, as enterprises get better at data collection via the Internet of Things and internally generated machine/customer data and the number of external data services providers grow, the data barrier to adoption is likely to become less daunting. Additionally, as the AI/machine learning skills gap widens, a combination of college graduates with the relevant skillset, re-training in AI/ML, AI/ML consulting firms, and better tools which automate the process are likely to emerge to fill the gap. The net of this, is that we believe most large enterprises (or smaller, data-centric enterprises) are likely to eventually at least experiment with machine/deep learning.

Due to the pace of innovation in the space, the technology landscape for developing a machine learning pipeline is still very fragmented. However, the emerging “AI stack” shares similarities with analytics in the mainframe, client-server, and current era analytics and development stacks. As highlighted in the “Evolution of the Stack” graphic below, the components of the stack ranging from the tools, to the languages, to the storage remain present.

The primary difference between the AI stack and prior technology shifts is that the bulk of machine learning pipelines heavily rely on open source technologies and services provided by cloud platform vendors. The drivers of this shift are multi-fold, but include 1) need for on-demand compute and storage to store and process large amounts of data, 2) heavy investment by cloud services providers such as Microsoft, Amazon and Google into machine learning services, and 3) the embrace of open source as a standard by large enterprise buyers in order to avoid vendor lock-in and reduce costs.

GPU COMPONENTS

GPUs and FPGAs emerging as key components in the AI stack

The repurposing of GPUs for deep learning has been one of the key drivers of our current “AI spring”. Within the AI/ML ecosystem, there are two primary applications that determine the performance of a neural network with each requiring a different resource setup. The first is the construction and use of a training algorithm. The training algorithm leverages a large (usually the larger, the better) data set to find correlations and build a model that can determine the probability of an output, given a new input. Training is very resource-intensive, and most modern training is done on GPU-powered systems.

The use of models and algorithms once they have been trained is referred to as inference. Inference requires far less computing power, and typically combs through smaller, incremental data input sets. Given the single-purpose nature of inference, specialized silicon is being developed specifically for that application, referred to as FPGAs (Field Programmable Gate Array) and ASICs (Application Specific Integrated Circuit). This type of integrated circuit was originally developed for prototyping CPUs, but is increasingly being used for inference in artificial intelligence. Google's Tensor Processing Unit is an example of an ASIC purpose-built for AI and machine learning. Microsoft has been using FPGA chips for inference as well. Intel acquired FPGA manufacturer, Altera, in 2015 on the view that by 2020, a third of data centers could be leveraging FPGAs for specialized use cases.

COMPUTE

Cloud platform vendors jostling for share

Given the cost-prohibitive nature of building true AI or ML capabilities in-house and on-premise, and the improving options available from public cloud providers, we believe relatively few enterprises will choose to build on-premise solutions. This creates an opening for providers such as Databricks (which offers Spark in the cloud as well as a number of tools to support the machine learning process), as well as from major cloud platform providers.

Offerings from major platform providers are comparable, but with some key differences that make one solution or another more applicable for specific use cases. While there are many vendors with GPU-based cloud offerings, we focus our analysis on those with the greatest ability to scale, and those that have been cited most frequently in our conversations with users.

STORAGE

Hadoop and Cloud Object Storage Emerging as Key Enabling Technologies

For deep learning in particular, large amounts of data improve the performance of machine learning models. Data growth in many industries has hit an inflection point. For example, in computational biology, the amount of usable data today is estimated by the Broad Institute to be north of 200 petabytes and growing faster than consumer web data. Petabyte scale data tends to be scored in one of two environments: Hadoop clusters (in HDFS) or in a cloud object storage service such as Amazon S3. Scale-out storage solutions from providers such as Dell's EMC division (e.g. Isilon) are also likely to be used in some environments. However, we believe open source or cloud-based storage services are likely to capture the bulk of incremental data created. This is primarily due to the low cost of these options versus on-premise, proprietary alternatives and the ability to flexibly scale up and scale-down usage in the cloud.

DATA MOVEMENT AND TRANSFORMATION

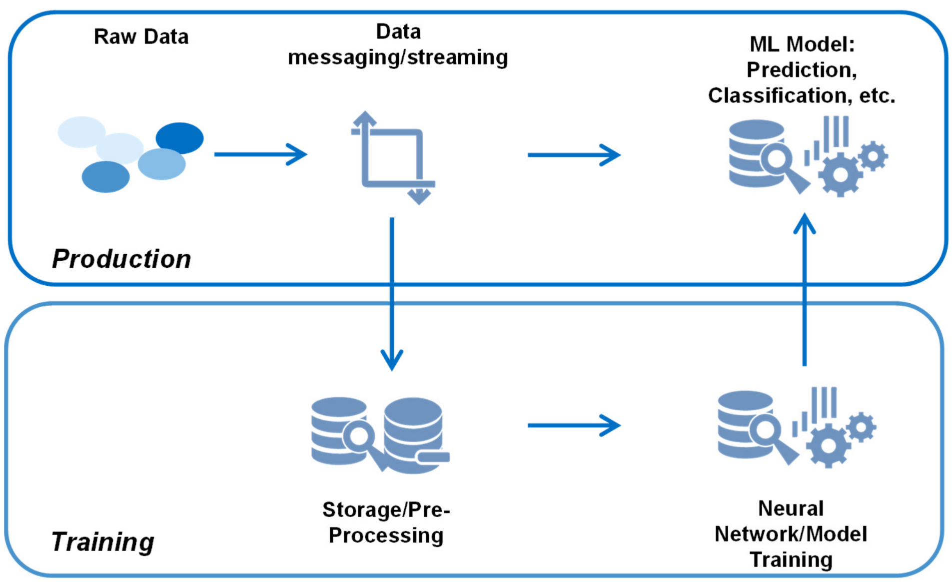
Kafka (supported by Confluent) and Spark (Databricks, Cloudera, Hortonworks) emerging as key technologies

Messaging, stream-processing and data transformation are key components to a machine learning pipeline. While a model is being trained, data is streamed into a storage system before being prepared and fed into a neural network or other machine learning framework. Once a model has been created, "live" data from sensors, the web, or other sources is streamed and prepared for analysis by the model, which then analyzes the data in real-time. Historically, ETL vendors (such as Informatica and IBM) and messaging vendors (such as TIBCO) have been providers of streaming and stream processing technology. Over the last five years, that has changed. In most of the machine learning environments we observed during our research, open source solutions such as Kafka, Storm, and Spark were utilized heavily. Messaging services such as Amazon Kinesis and Google Pub/Sub were also used.

Even for neural networks, data needs to be prepared. For example, images and text are normalized to the same size and color (images) or format (text). For these tasks, custom code can be written or tools such as Skyminds's DataVec can be used.

Exhibit 16: Machine Learning in Production

How various open source and cloud technologies would be utilized in the machine learning pipeline



Source: Company data, Goldman Sachs Global Investment Research

DATA PROCESSING

Neural Networks, Spark come to the forefront

The database/data processing market has historically been one of the largest and most lucrative in software. In 2015, for example, Gartner estimated the size of the database market at $35.9bn. One of the largest companies in the S&P (ORCL, with a >$160bn market capitalization) derives the bulk of its profits from its database product. In AI, a new set of technologies are being utilized. First, the neural network has emerged as a key data processing technology. As we explained in the "What is AI" section, neural networks processing inputted data via nodes to create outputs. For example, inputs might be emails or images and the outputs might be "spam" or "cat". To date, the creation of neural networks has mostly been done via custom development using a variety of frameworks (such as Google TensorFlow or Caffe). Cloud services such as Google Cloud Machine Learning are also emerging to enable developers and data scientists to build neural networks in the cloud.

The usage of Spark as a processing technology was a common theme in our discussions with VCs and companies. Spark remains one of the fastest growing open source projects (currently with over 10k Github stars). and has received heavy investment from IBM, Cloudera, Hortonworks and Databricks (which has the bulk of the committers to the project).

PROGRAMMING LANGUAGES

Python, the language of choice for machine learning

AI and machine learning is still in early stages of adoption. This means that custom development remains the primary avenue for creating production applications and workflows. The languages of machine learning and data science are Python and R. Python has not been monetized to date. In the R ecosystem, Microsoft (which acquired Revolution Analytics) and RStudio (an open source provider) are the primary vendors.

Exhibit 17: Key Open Source Projects in the Machine Learning Pipeline
Project, supporting company, and venture funding where applicable

Open source project name	Description	Major supporting vendors	Funding ($ in mn)
Hadoop	Framework that allows for distributed storage and processing of large data sets on clusters built from commodity hardware	Cloudera, MapR, Hortonworks (HDP)	Cloudera $1,004 MapR $194
Spark	In-memory data processing engine	Cloudera, Databricks, Hortonworks, MapR	Databricks $47
Kafka	Message broker project in Scala for handling real-time data feeds	Confluent, Cloudera, Hortonworks	Confluent $31
Deeplearning4j	Deep learning library in Java and Scala	Skymind	$3
Storm	Real time data processing system	Hortonworks, MapR	See "Hadoop" above
Theano	Numerical computation library for Python	NA	NA
Caffe	Deep learning framework based in C++	NA	NA
TensorFlow	Machine learning library for numeral computation using data flow graphs	Google	NA
Torch	Scientific computing framework based in the scripting language LuaJIT and underlying C/CUDA implementation	NA	NA

Source: Company data, Project websites, TechCrunch

ANALYTICS TOOLS

Multiple vendors seeking to improve data scientist productivity

Throughout the history of analytics, tools have emerged to enable businesses to extract value from data without relying on custom development. Advanced statistics tools such as SAS Institute and SPSS, BI solutions such as Microstrategy and Business Objects, reporting solutions such as Crystal Reports, and, more recently, data visualization providers such as Tableau have monetized the need to improve the productivity of the business analysts and data scientists who support business users.

Machine learning tooling is beginning to emerge to accelerate the productivity of data scientists. An example is Microsoft's Azure Machine Learning solution, which creates a drag and drop interface for data scientists to create machine learning workflows. Data scientist focused tools from SAS also provide tooling to enable the development and deployment of various machine learning libraries.

Consulting Services: Monetizing the skills gap

As we have noted previously in this report, talent remains one of the primary barriers to machine learning adoption. Applied machine learning also provides an opportunity for legacy technology providers with large consulting businesses to more effectively monetize open source technologies (via wrapping them in consulting solutions).

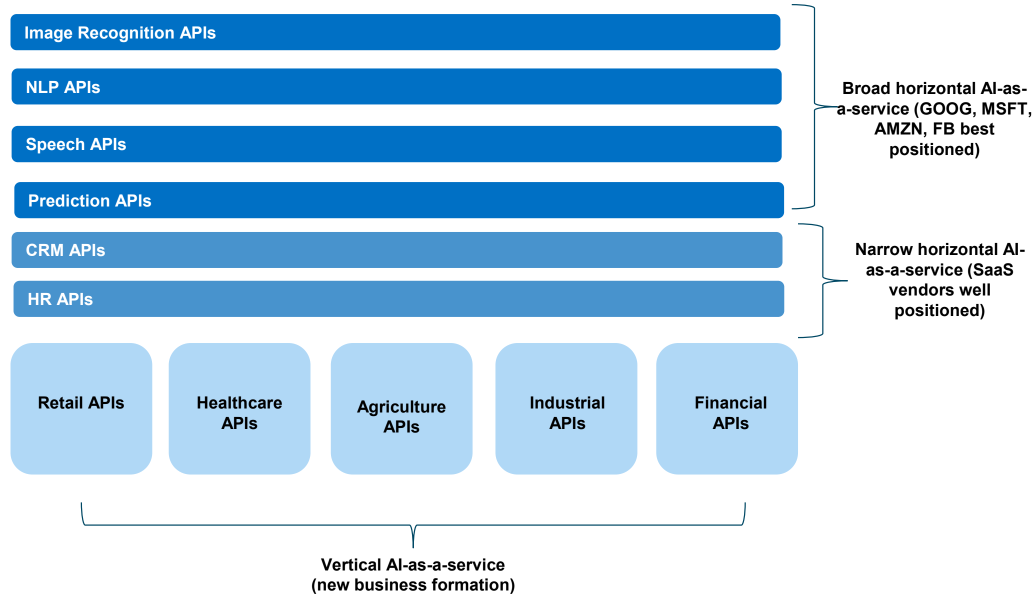
Other business models are also emerging to close the skills gap. Kaggle, as an example, crowdsources machine learning via hosted competitions. Data scientists can win prize money, practice on "real-world" datasets, and build a machine learning portfolio. Businesses are able to access talent to solve problems without having to invest heavily in a machine learning team.

AI-aaS: Likely the biggest driver of new market creation

While we expect many companies will invest in DIY AI, creating growth for picks and shovels providers, we see the most potential for dynamism and new business creation in AI-aaS. Because large, unique datasets are relatively limited and scarce AI talent is likely to consolidate to such datasets, in our view it seems unlikely that a large number of enterprises are building their own neural networks in five years. We believe a more likely scenario is that a large number of AI services providers emerge that 1) have access to unique data sets and 2) because of access to unique data sets, also attract talent necessary to create value-added AI services.

AI-aaS offerings are most commonly delivered via an API. The most basic example is a developer who wants to add image recognition capabilities to her app. Rather than acquiring a large data set of images and training a model, the developer instead accesses an Image API via a horizontal AI-aaS provider such as Clarifai, Google or Microsoft. When speech recognition is utilized in the app a call is made to the API in the cloud and the image is classified by a trained machine learning model.

Exhibit 18: AI-as-a-Service (AI-aaS) landscape
Machine learning APIs are developing to address horizontal and vertical use cases



Source: Company data, Goldman Sachs Global Investment Research

We see the market for AI-aaS evolving along at least three fronts, as highlighted below and in Exhibit 18 above.

Broad horizontal AI-aaS (images, voice, text, etc.)

Both Google and Microsoft offer APIs for speech, translation, and image recognition for as little as $0.25 per 1,000 API calls/month. Developers can utilize these APIs to embed AI capabilities into their applications. For core horizontal AI such as NLP and image recognition, we view large cloud platform providers as best positioned due to their possession of large data sets enabling more accurate AI services and the ability to refine their results over billions of uses by actual consumers.

Exhibit 19: Horizontal AI-aaS Offerings and Pricing
Sample of AI-aaS offerings from cloud platforms

Company	Product	Description	Pricing (US)
AMZN	Machine Learning API	Managed service for generating ML models and predictions; includes modeling APIs and batch/real-time prediction APIs	Data analysis and model building $0.42 per compute hour; Batch Predictions $0.10 per 1,000 predictions; Real-time $0.0001 per prediction
GOOGL	Vision API	Image analytics tool	Free to $5 per 1k units depending on features used and monthly usage
	Google Cloud Machine Learning	Managed services enables users to build machine learning models	Training clusters: $0.49/hour to $36.75/hour depending on training units per hour Prediction requests: $0.05 to $0.10 per 1k requests + $0.40 per node hour depending on number of requests
	Speech API	Converts audio to text	0-60 minutes free; 61-1mn minutes $0.006 per 15 seconds
	Cloud Natural Language API	Enables analytics of unstructured text	0-5k units free; above 5k pricing ranges from $0.125 to $1 per 1k units depending on features used and monthly usage.
	Translate API	Translates and detects 90+ languages	$20 per mn characters
	Prediction API	ML/predictive analytics tool	Limited free use for 6mo; paid usage $10/mo/project access fee, free predictions and streaming training up to 10,000 per day, additional predictions at $0.50 per 1k predictions, additional streaming updates at $0.05 per 1k updates. Training data $0.002 per MB.
MSFT	Computer Vision API	Visual data analytics tool	Free to $1.50 per 1k transactions depending on monthly usage
	Emotion API	Detects emotions in images	Free to $0.25 per 1k transactions depending on usage; free for video
	Face API	Enables face detection with attributes and face recognitio	Free to $1.50 per 1k transactions depending on monthly usage
	Text Analytics API	Enables analytics of unstructured text	Free to $2,500 per month depending on usage
	Video API	Advanced algorithms for tracking faces, detecting motion, stabilizing and creating thumbnails from video	Free; 300 transactions per month per feature
	Bing Speech API	Coverts speech to text and back to speech, enabling app to "talk back" to users	Free to $4 per 1k transactions or $5.5-$9 per hour depending on type and usage
	Custom Recognition Intelligence Service	Customized speech recognition tool	Private preview by invitation only
	Speaker Recognition API	Enables identification of speakers and speech as a means of authentication	Free to $10 per 1k transactions depending on usage and features used
	Bing Spell check API	Contextual spell checking	Free to $450/month and overage at $50 per 100k transactions depending on per month usage
	Language Understanding Intelligent Service (LUIS)	Teachs apps to understand commands from users	Free to $0.75 per 1k transactions depending on usage
	Linguistic Analysis API	Natural language processing tools that identify structure of text	Free; 5k transactions per month, 2 per second
	Web Language Model API	REST-based cloud service providing tools for natural language processing	Free to $0.05 per 1k transactions depending on usage
	Academic Knowledge API	Interprets user queries for academic intent and retrieves information from the Microsoft Academic Graph	Free to $0.25 per 1k transactions depending on usage
	Entity Linking Intelligent Services	Contextualized language processing	Free trial; 1k transactions per day
	Recommendations API	Generates personalized product recommendations	Free to $5,000 per month depending on usage
	Bing Autosuggest API	Sends a partial search query to Bing and gets back a list of suggested queries	Free to $270/month and overage at $30 per 100k transactions depending on per month usage
	Bing News/Image/Video/Web Search API	Sends a search query to Bing and gets back a list of relevant results	Free to $8,100/month and overage at $30 per 10k transactions depending on per month usage

Source: Company data

Narrow horizontal AI-aaS (customer churn, employee retention, etc.)

For more focused horizontals such as CRM (lead scoring), HR (talent retention), and manufacturing (predictive maintenance) we see SaaS vendors as well positioned, due to the large amounts of differentiated data that SaaS vendors have access to. Workday, Salesforce.com, Zendesk, Oracle, SAP and IBM are vendors who could eventually compete for narrow AI-aaS use cases. Most SaaS vendors we have spoken with are investing in data benchmarking and analytics products, with the view that their data is a barrier to entry longer-term.

Vertical specific AI-aaS (medical imaging, fraud prediction, weather prediction, etc.)

Vertical specific AI-as-a-service is likely to drive more diversity. Large industry giants could aggregate data, build machine learning models, and sell access to the model to partners, customers, and suppliers. Start-ups can build unique datasets in use-case specific verticals such as medical imaging and enable hospital networks to access APIs. Industry consortiums in areas such as retail or advertising could pool data to better compete against larger competitors (e.g., retailers could pool data to better compete against Amazon's recommendation engine).

Exhibit 20: The Vertical AI-aaS Landscape in Healthcare

Company name	Description	Funding ($ in mn)
AiCure	Provides AI-based technology that visually confirms medication ingestion on mobile	$12
Apixio	SaaS application that increase Medicare Advantage risk adjustment accuracy, productivity and speed	$42
Arterys	SaaS analytics platform for medical imaging with big data quantification and visualization capabilities	$14
Atomwise	Develops AI systems for drug discovery	$6
BioBeats	Creates wellness solutions based on AI/machine learning	$3
Butterfly Network	Develops high performance ultrasound machine with deep learning algorithms	$100
Deep Genomics	Develops machine learning based technologies for precision medicine, genetic testing, diagnostics, and therapy development	$4
Enlitic	Uses deep learning to generate actionable insights from clinical cases for doctors	$15
Entopsis	Diagnostics solution focused on oncology, autoimmune disorders and rare diseases	$1
Ginger.io	Analyzes behavioral data produced by smartphones for mental-health monitoring	$28
Healint	Chronic disease management solution	$1
Lumiata	AI-powered health data analytics tool	$20
MedAware	Prescription error prevention solution	$2
Numerate	Machine learning based drug design platform	$17
Oncora Medical	Clinical decision support software for oncologists leveraging big data analytics and machine learning	$1
VisExcell	Develops computer-aided detection in mammograms and other imaging modalities through big data and machine learning algorithms	NA
Wellframe	Patient management and analytics solution for care managers	$10
Welltok	Provider of machine learning-based health management tool that organizes health programs, communities, apps and tracking devices	$164
Zebra Medical Vision	Provides automated analysis of real-time and retrospective medical images	$20
Zephyr Health	Insights-as-a-services solution for Life Sciences companies	$33

Source: Company data, Crunchbase

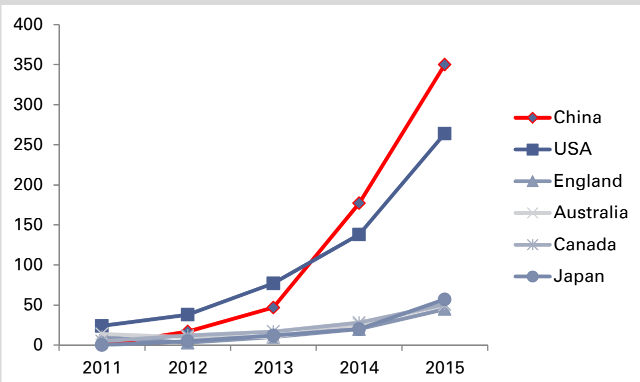
China and the state of AI

The country's AI market is estimated to grow to Rmb9.1bn in 2020 from Rmb1.2bn in 2015, according to iResearch. In 2015, roughly Rmb1.4bn (+76% yoy) capital flowed into Artificial Intelligence (AI) in China.

In terms of government policy, China's National Development and Reform Commission (NDRC), together with other relevant government bodies, published "*Three-year Implementation Plan for "Internet Plus" Artificial Intelligence*" on May 18, 2016. The implementation plan laid out six specific areas of support for AI development, including capital funding, system standardization, IP protection, human capital development, international cooperation and implementation arrangements. The plan targets the establishment of basic infrastructure and innovation platform, industry system, innovative service system and basic industry standardization of AI in China by 2018. NDRC expects the overall China AI industry to be synchronized with international development, and lead the global market in system-level AI technology and applications.

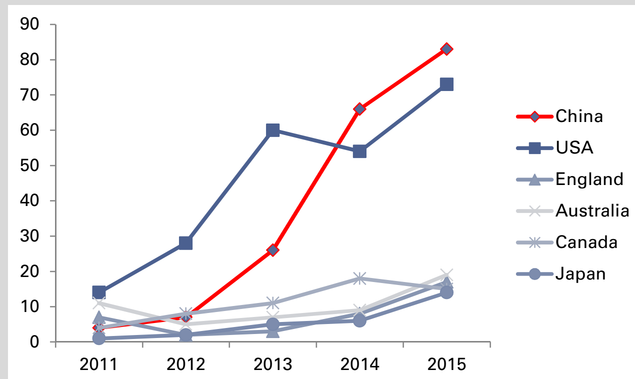
China has made major moves, and based on the number of journal articles cited mentioning "deep learning" or "deep neural networks", China surpassed the US in 2014 (Exhibit 23). The AI research capability in China is also impressive (Exhibit 24), as it has the world-leading voice and visual recognition technology. Deep Speech 2 developed by Baidu in Nov 2015 was able to achieve 97% accuracy, being recognized as one of the top 10 breakthrough technology in 2016 by MIT Tech Review. In addition, DeepID developed by Chinese University of Hong Kong achieved 99.15% face recognition accuracy in LFW (Labelled Faces in the Wild) as early as 2014.

Exhibit 21: Journal articles mentioning "deep learning" or "deep neural network"



Source: US National Science and Technology Council

Exhibit 22: Journal articles cited at least once, mentioning "deep learning" or "deep neural network"



Source: US National Science and Technology Council

At present, the field of AI in China covers:

1) Basic services such as data resources and computing platforms;

2) Hardware products such as industrial robots and service robots;

3) Intelligent services such as intelligent customer services and business intelligence; and

4) Technical capabilities such as visual recognition and machine learning.

According to iResearch, voice and visual recognition currently contributes 60% and 12.5% of the total China AI market, respectively. Of the AI-related companies in China, 71% are focused on developing applications. The rest focus on algorithm, 55% are looking into computer vision, 13% on natural language processing, and 9% on fundamental machine learning.

In our view, the key players at the forefront of artificial intelligence are likely to continue to be in the US and China.

Bots: The future of user interface

Bots are potentially paradigm-shifting. In a bot-centric world, the user experience evolves from click-based to conversational (text or voice) and interaction shifts from web or app-oriented to messaging or voice platform-oriented. In other words, rather than opening three different apps to book travel, shop for clothes, and engage customer service, a user could instead engage in conversation with assistance-providing bots via a messenger, simultaneously. As a result, we see wide-ranging implications across e-commerce, customer support, and employee workflows and productivity.

A key driver over the last 12-18 months has been large cloud and internet companies creating and open-sourcing machine learning frameworks. In late 2015 Google open-sourced TensorFlow, a library of machine learning algorithms and Amazon and Microsoft have also been active, releasing cloud services to support machine learning projects of their own. We expect this trend toward democratizing machine learning will continue to spur development of intelligent bots as major players (Amazon, Google, Apple, Microsoft) look to integrate conversational interfaces (Alexa, Google Assistant, Siri, Cortana) throughout their respective ecosystems. Following Samsung's acquisition of Viv this year, we expect further integration of the Viv AI-based digital assistant in Samsung's ecosystem of devices and smartphones as well.

Natural Language Processing (NLP) is the field of AI broadly encompassing interactions between human and computer languages

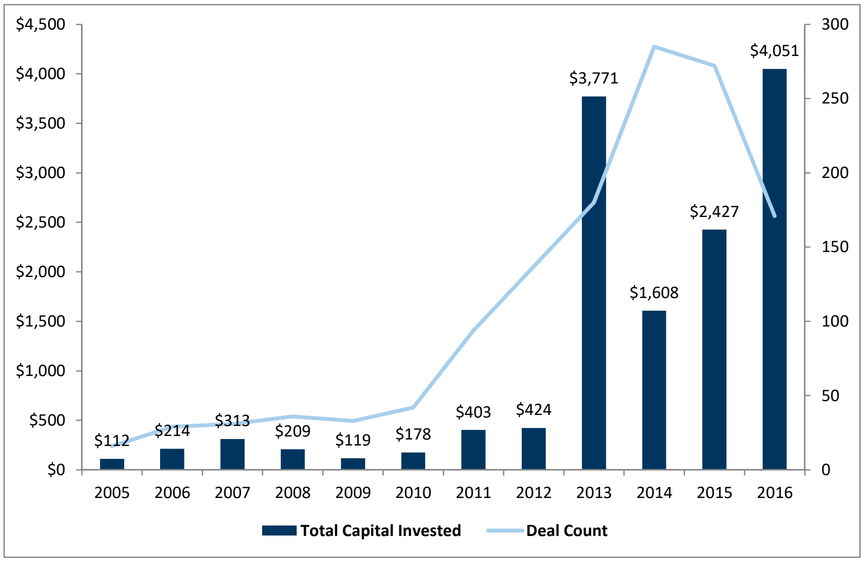
Natural language processing (NLP). The promise of bots is rooted in their potential to be intelligent and process natural language. Accordingly, the rise of interest in bots has coincided with a rise of interest and innovation in machine learning, the tech underlying the AI field of Natural Language Processing (NLP), or computer understanding, manipulation, and derivation of meaning from language. In contrast to word processors that operate much like the CTRL+F function built around hard-coded rule sets, NLP leverages machine learning algorithms to learn rules based on large sets of training data that can then be applied to new sets of text. The core tenet of machine learning applies in that the more data the NLP system ingests, the more accurate and broader its applications become.

While early applications of NLP have been seen in text mining (e.g., analysis of legal documents, insurance policies, and social media) and automated question answering, advances in neural networks and deep learning models are allowing NLP systems to become increasingly intelligent and manage the ambiguity that plagues human language. Google's open-source foundation for NLP implemented in TensorFlow, SyntaxNet, leverages neural networks to combat ambiguity in left-to-right processing by discarding potential hypotheses only once several other higher ranking ones are discovered. As a result, SyntaxNet models are some of the most complex ever trained within the TensorFlow framework, according to Google.

Messaging platforms. The rise of bots has coincided with the rapid growth of messaging applications such as Facebook Messenger, WhatsApp and, in the enterprise, Slack and HipChat. Messaging applications provide a medium through which bots can interact with users across iOS, Android and the web. Further, larger messaging applications are evolving into platforms which support multiple interaction types. On Slack, an enterprise user might collaborate with a team, monitor an application, create a to-do list or monitor expenses from the same interface. On Facebook Messenger, a user may chat with friends, raise a support issue with a brand, or order an Uber from the same interface.

Recent chat bot acquisitions or partial acqui-hires by Amazon (Angel.ai CEO) and Google (API.ai), each specializing in conversational interface technologies, highlights the opportunity both companies and investors see in the union of chat and natural language process capabilities. Since 2013, roughly $12bn in cumulative VC funding has been invested across private messaging companies within AI/ML, ecommerce, SaaS, and Cybersecurity verticals vs. approximately $2bn in the 8 years prior, according to PitchBook.

Exhibit 23: VC funding across messaging
$ millions



Source: PitchBook

Some beneficiaries are emerging in the event bots proliferate. The first group of beneficiaries is the messaging platforms. Bots drive increased engagement and create opportunities to drive commerce on such platforms. The second group is the hardware and infrastructure providers, which range from GPU providers to open source vendors, data platform vendors, and cloud services providers. Others who are tapping into bot capabilities include software providers, who see bots as a potential means for automating enterprise customer service.

Digital Personal Assistants. Many companies have been utilizing complex algorithms, machine learning, and big data software to create recommendation engines based on analysis of past behavior and customer data for some time. These engines are being employed in an effort to influence purchase behavior, but much of the same technology is utilized in the engineering of Digital Personal Assistants, or bots with the ability to complete or automate simple tasks based on voice commands.

Merging the complex forecasting and inference capabilities of recommendation engines with voice recognition software has led to the creation of Apple's Siri, Amazon's Alexa, Google Assistant, and Microsoft's Cortana. Leveraging machine learning and cloud infrastructure, these applications improve as they gather more information about the user: speech patterns, interests, demographics, spending habits, schedule, occupation, likes, and dislikes. Most, if not all, of this information can usually be gleaned by software monitoring a person's use of a smartphone or connected device (Amazon Echo, Google Home). As these Digital Personal Assistants access more data, the analytics should allow them to differentiate between similar requests from different users, becoming increasingly personalized. For example, the instructions "show me the best camera" could mean different things to different consumers. A powerful analytics engine coupled with user data can help determine whether the user would prefer the least expensive camera, the most highly reviewed one, or some other combination of traits equating to "best" for that individual.

We see continued innovation in data aggregation and analytics driving improvements in AI-powered Digital Personal Assistants. We also expect serial innovators like Amazon and

Google to continue to remove friction points in purchase process (Echo, Echo Dot) and further engrain themselves in daily tasks (Google Home).

USE CASES

Agriculture

$20bn total addressable market by 2025

We believe machine learning has the potential to increase crop yields and decrease fertilizer and irrigation costs, while assisting in earlier detection of crop/livestock disease, lowering labor costs associated with post-harvest sortation, and improving quality of produce and protein that comes to market. As we see a proliferation of sensors used to gather soil, weather, aerial/satellite imagery, and even auditory data, we believe that the insights generated from deep learning algorithms on those petabyte-scale data sets will inform (and sometimes make) decisions regarding planting times, irrigation, fertilization, and livestock care, and lead to an increase in land, equipment and human productivity across the spectrum of agriculture. **Given that all of the identified technologies used in digital agriculture would be either optimized, or completely powered by ML / AI, we assume that 25% of that value creation accrues to vendors in the ML / AI chain, which would imply a TAM of $60bn within the $1.2tn crop agriculture market by 2050. Assuming linear adoption over that timeframe implies a roughly $20bn TAM by 2025.**

What is the opportunity?

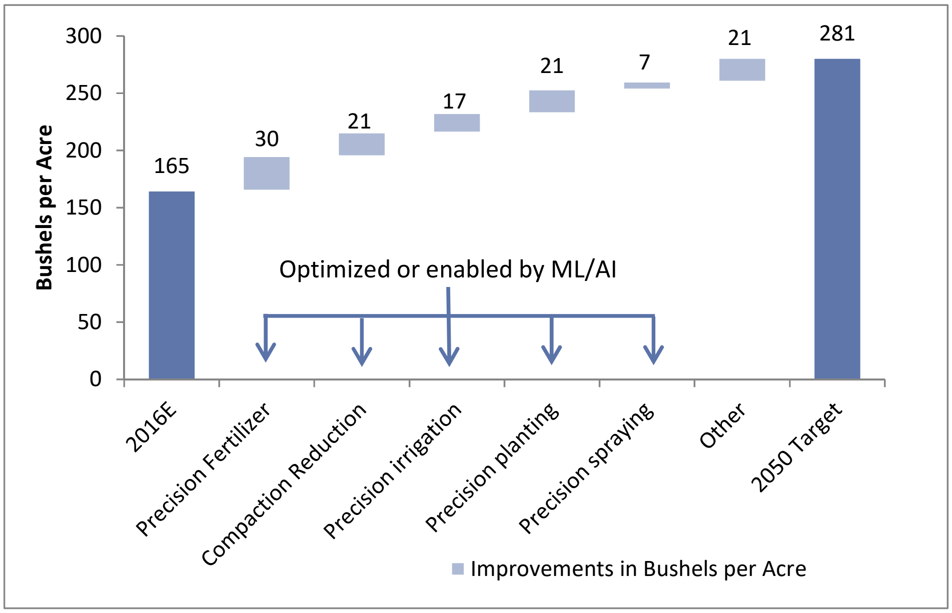
There is meaningful production and yield loss as well as labor expense that can be reduced through machine learning applications in agriculture. Just in US corn production, our equity research teams have identified technologies ranging from precision fertilizer to compaction reduction that they believe could improve corn yields 70% by 2050. Importantly, every innovation identified in their research is powered and enabled by machine learning and artificial intelligence.

We have identified several specific areas within agriculture where we see particular benefit from the application of ML and AI technologies. The Farmers Business Network, for example, is an organization that aggregates data on seed performance, agronomic practices, input prices, yield benchmarking, and other farmer-submitted data, in an effort to utilize deep analytics to improve yields.

Utilizing sensor, weather, soil, and even drone/satellite imagery data, machine learning can help determine best practices based on current and expected weather patterns, impact from crop rotations on soil quality, and help farmers optimize fertilization, irrigation, and other decisions. Analysis of aerial imagery can potentially help identify crop disease such as soybean rust more quickly and efficiently than human observation and early treatment can prevent loss of harvest.

The same pattern recognition technology can be used to identify disease and lameness (infection or injury in legs/feet/hooves that impacts mobility and overall wellness) in livestock animals as well. Lastly, we see applications for replacing human inspectors along grading and sortation lines for product and meat products using visual imagery and automated sortation facilities.

Exhibit 24: All of the innovations identified for improving crop yields involve applications of machine learning and artificial intelligence
US corn yield drivers



Source: Goldman Sachs Global Investment Research, USDA, company data

What are the pain points?

Crop yields depressed by sub-optimal fertilization, irrigation, and pesticide controls. In the GS research report, *"Precision Farming: Cheating Malthus with Digital Agriculture"* (July 13, 2016), several problems were identified that could be solved by gathering the proper data and performing the proper analytics. This is critically important as feeding the projected world population in 2050 requires a 70% in increased crop yields.

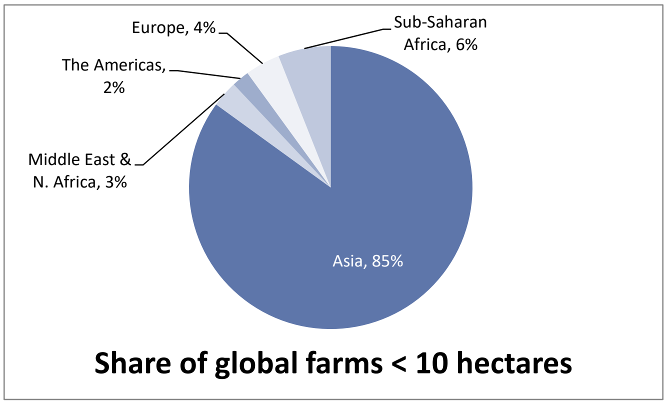
Increasing labor costs. Agriculture has historically turned to technological innovation to offset labor costs, and we believe machine learning is the next step in that evolution, particularly within post-harvest/slaughter sorting where much of the visual inspection of produce and meat products is still done by human workers. According to the Bureau of Labor Statistics, 53k individuals are employed in the United States as "graders and sorters, agricultural products", generating roughly $1.3bn in annual labor costs. According to BLS data "pesticide handlers, sprayers, and applicators" represent another $1.3bn in labor costs within agriculture.

Losses due to animal disease/distress. We estimate over $11bn in annual loss within global dairy farming due to preventable lameness among dairy cows. Academic research indicates that between milk loss, decreased fertility, and treatment costs, lameness costs dairy farmers ~$175 per instance and occurs at a rate of 23.5 instances per 100 cows per year, which implies over $11bn in costs when applied to the ~250mn dairy cows globally.

What is the current way of doing business?

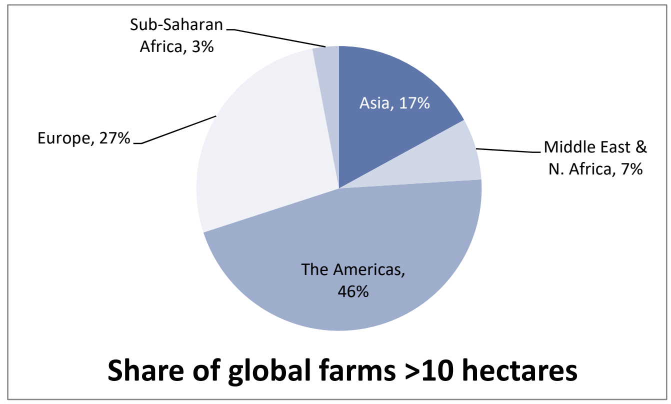
The vast majority of farms are small, but the majority of agricultural land is controlled by large farms. According to the UN's FAO, 72% of all farms globally are less than 1 hectare in area, and while only 1% of all farms are larger than 50 hectares, those large farms control 65% of global agricultural land. Farms over 10 hectares are overwhelmingly found in more developed geographies like the Americas and Europe (73% between the two), while Asia accounts for 85% of farms smaller than 10 hectares. As such, most of the world's farmland has access to infrastructure and economic development that would enable the use of precision farming techniques, so long as those techniques are financially viable solutions.

Exhibit 25: Small farms are the norm in the developing world…



Source: FAO; Goldman Sachs Global Investment Research

Exhibit 26: While the developed world has seen consolidation and scale of farming operations



Source: FAO; Goldman Sachs Global Investment Research

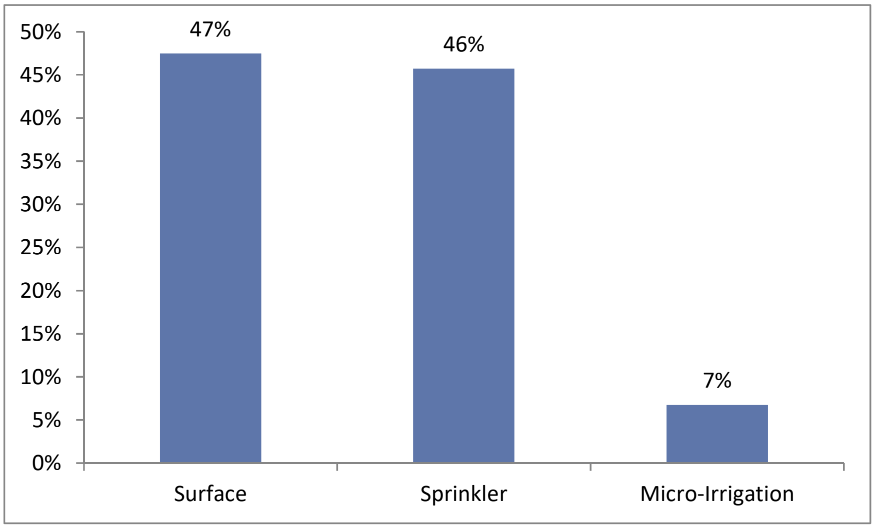
Even within economically developed countries, precision agriculture remains in early stages. Irrigation for example, is still carried out via flooding or other forms of surface irrigation, which remains one of the least efficient and least technologically advanced methods. For the major areas of crop cultivation, current technologies include:

- Fertilizer: weather and field monitoring, and blanket application.
- Planting: multi-seed planters, variable rate planting, and crop rotation.
- Pesticides/herbicides: satellite and drone imagery already in use in some larger-scale operations for target areas. Smaller operations using blanket applications.
- Irrigation: flooding and other surface irrigation, central pivot sprinklers, drip systems, and hybrid sprinkler/drip systems.
- Harvesting/sorting: much of the harvest for crops like corn and wheat has already been mechanized on large farms. Some sorting has been automated (by size and color).

We are also seeing the advent of a democratization of data in farming, with the establishment of the Farmers' Business Network (FBN) in the United States. FBN is an independent business to which farmers can subscribe and submit farm data that is then anonymized. Aggregate farm data is used by FBN in analytics processes to generate insights for individual member farmers leveraging yield, time, weather, and other data.

Within livestock and dairy farming, current technologies include universal application of antibiotics or other preventative medicines, vaccinations, culling of sick animals, and chemically balanced feed supplements. In addition, cattle operations have also employed foot baths to prevent and treat hoof illness and infections.

Exhibit 27: Within the US, nearly half of all irrigated agricultural land is irrigated via flooding or other surface irrigation- one of the least efficient and least technologically advanced methods
% of US irrigated agricultural land, by irrigation method.



Source: FAO, Goldman Sachs Global Investment Research

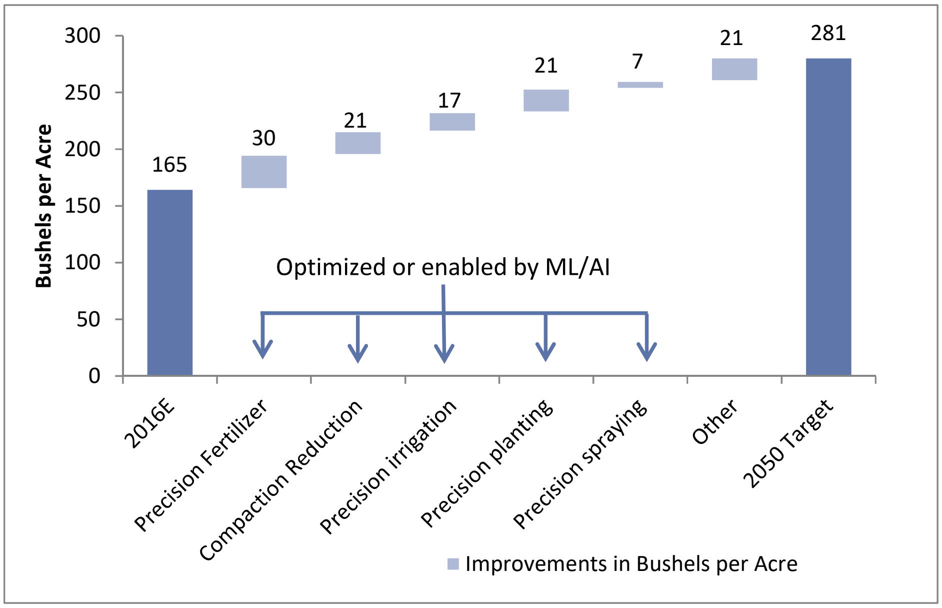
How does AI/ML help?

Machine learning's ability to use large data sets to optimize around a single or set of end goals lends itself favorably to solving issues in agriculture like crop yields, disease prevention, and cost efficiency.

In post-harvest sorting and pesticide applications we believe that ML /AI can reduce costs and improve efficiencies to create $3bn in annual labor savings over time, just within the United States. Global numbers would likely be more than double that figure, by our estimates. Finally, we believe that ML / AI applications can improve breeding and health conditions, leading to roughly $11bn in value creation (between recouping lost potential revenue and absolute cost reduction) in dairy farming, and $2bn in poultry production, just from two common maladies impacting animals.

Improving crop yields. People are already utilizing nearly all of the planets useable agriculture land, and the United Nations expects global population to reach 9.7bn by 2050. As such there is a need to improve crop yields in order to meet the future global demand for food. Machine learning can be used to analyze data from drone and satellite imagery, weather patterns, soil samples, and moisture sensors to help determine optimal method of seed planting, fertilization, irrigation, spraying, and harvesting.

Exhibit 28: Machine learning plays an important role in nearly every innovation identified in our Precision Farming report (published on July 13, 2016)
Potential improvement in corn crop yields, by technology

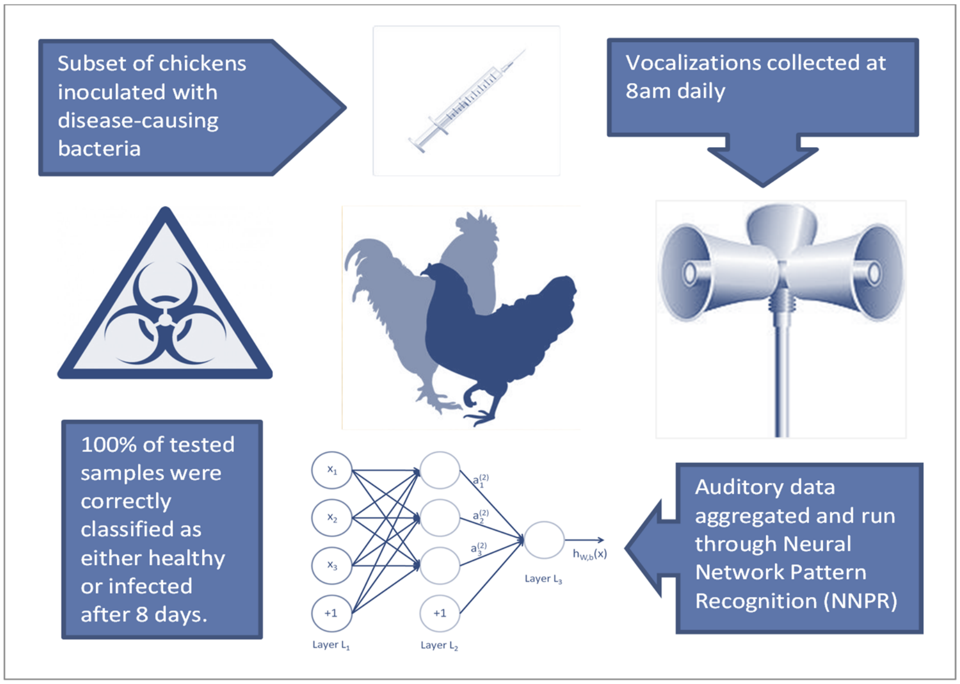


Source: Goldman Sachs Global Investment Research, USDA, company data

Post-harvest sorting labor. We see a simple case study in a Japanese cucumber farmer's application of Google's TensorFlow ML technology to automate the process of sorting his cucumbers – a process that had historically required significant manual/visual inspection and labor costs. Using simple, inexpensive hardware including Raspberry Pi processors and ordinary webcams the farmer was able to utilize TensorFlow to train an algorithm that could sort the cucumbers into 9 categories with a relatively high degree of accuracy, virtually eliminating the labor cost associated with sorting. We believe similar applications could be scaled much larger and used for agricultural products with high sortation needs and costs like tomatoes and potatoes.

Illness detection among poultry populations. In an academic research study, researchers collected and analyzed the sound files of chickens under the hypothesis that their vocalizations would change if they were sick or distressed. After collecting data and training a Neural Network Pattern Recognition algorithm, the researchers were able to correctly identify chickens infected with one of the two most common loss-causing diseases with 66% accuracy after 2 days of sickness, and with 100% accuracy after 8 days of sickness. Correctly diagnosing animals early enough to treat before loss occurs could eliminate the $2bn losses that industry experts estimate are caused by the disease.

Exhibit 29: Experiments have indicated that machine learning can correctly identify otherwise undetectable disease via auditory data analysis, virtually eliminating loss due to certain curable diseases



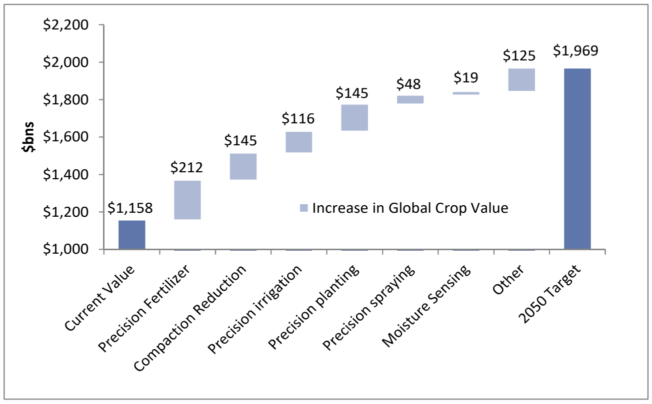
Source: An Intelligent Procedure for the Detection and Classification of Chickens Infected by Clostridium Perfringens Based on their Vocalization, Rev. Bras. Cienc. Avic. Vol 17. No. 4 Campinas Oct./Dec. 2015.

Quantifying the opportunity

Based on the potential increase in yields, crop input cost savings, dairy/livestock cost savings, sortation and labor savings, we believe machine learning could be used in processes that could generate over $1tn in value.

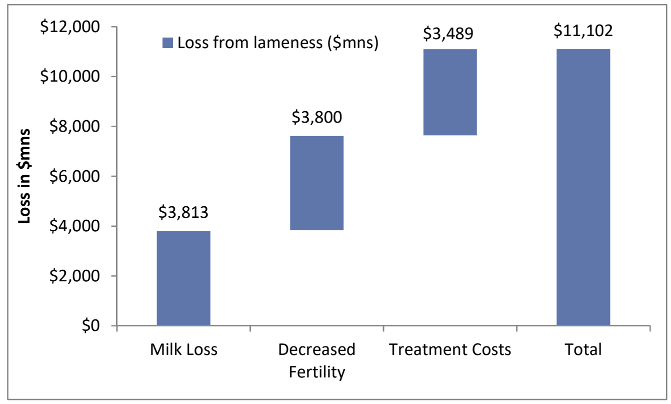
Within farming, we believe ML / AI can help drive up to 70% increase in crop yields. In Jerry Revich's Precision Farming note (*"Precision Farming: Cheating Malthus with Digital Agriculture,"* published on July 13, 2016), the TAM for digital agriculture was identified at $240bn, assuming 30% value accruals to the various technology vendors. Given that all of the identified technologies used in digital agriculture would be either optimized, or completely powered by ML / AI, we assume that 25% of that value creation accrues to vendors in the ML / AI chain, which would imply a TAM of $60bn in crop farming applications of ML / AI. Within protein agriculture, we believe that applications of machine learning (precision breeding mechanisms, disease prevention/treatment) could generate another $20bn.

Exhibit 30: Potential increases in global crop yields from advanced technologies could generate over $800bn in increased value within crop agriculture
Est. increases in global crop value, by technology ($, mn)



Source: Goldman Sachs Global Investment Research, company data

Exhibit 31: AI/ML can potentially eliminate over $11bn in losses to dairy farmers from the impact of just one type of cattle illness
Est. losses from lameness in dairy cows ($, mn)



Source: The cost of different types of lameness in dairy cows calculated by dynamic programming, College of Veterinary Medicine, Cornell University, NY; FAO; Goldman Sachs Global Investment Research

Who could be disrupted?

We believe machine learning has the potential to expand the global supply for crops, dairy, and livestock at a lower per-unit cost, based on cost savings from irrigation, fertilizer, labor, and disease prevention/treatment. We would expect disruption of the global market for fertilizer and pesticides/herbicides/fungicides, as well as veterinary pharmaceuticals as machine learning applications limit waste and improve preventative methods (limiting need for curative methods) in agriculture. We believe that most of this disruption is longer-term (5+ years), as we are still in early development stages for many of these technologies and the cost to early adopters is sometimes prohibitive, relative to other potential improvement mechanisms.

Farmers Business Network

We spoke with Co-Founder and CEO, Amol Deshpande, as well as members of the FBN engineering team. FBN is a network of 2,800+ member farmers, covering 10mn+ acres of farmland, that aggregates data uploaded from farmers and farm equipment in an effort to democratize that farm data and empower member farmers to leverage it for input pricing, seed selection, and yield optimization.

The Problem

Asymmetric information within farming has led to farmers making decisions about seed selection, fertilizer selection/application, and other essential aspects of their business without broad understanding of which options have created the highest yields in their region in recent years, or even whether the prices offered are comparable to those seen by other farmers. Farmers have also been subject to vendor biases.

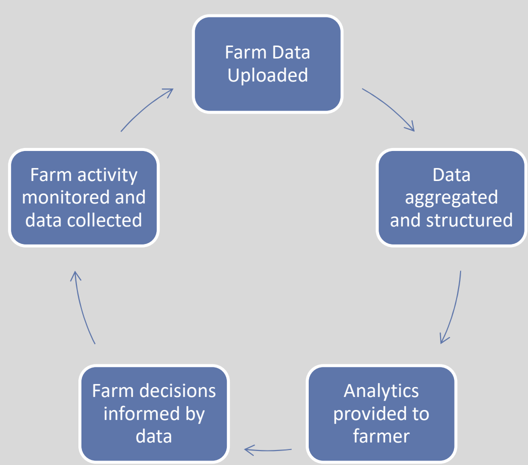
FBN Solution

Data Aggregation and Analytics: Farmers pay $500 annually (with multi-year discounts) to gain membership in FBN. Farmers can then upload data from their equipment and systems, including seed type data, soil data, harvest/yield data, as well as geolocation and elevation data, etc. FBN also aggregates data from other public sources including government and weather data. FBN uses machine learning to parse, cleanse, and analyze data in order to provide insights to individual farmers, tailored to individual farms, to help them select optimal inputs and farm management strategies to maximize yield and productivity.

Financing: FBN has also begun trials on captive finance services, with historical and predicted data from actual farms being used to determine creditworthiness. Without running credit checks, FBN was able to achieve 97%+ repayment rates, with the remainder on repayment plans.

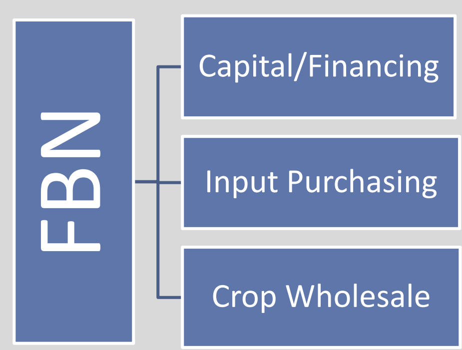
Procurement: Beginning with ag chemicals, FBN has begun offering procurement services to network farmers. Given FBN's access to massive input pricing data sets and its ability to purchase at larger scale on behalf of thousands of farmers, the company believes it can negotiate better pricing and drive down input costs while taking a 9%+ commission on each transaction. Average ticket size in early months has been $45k.

Exhibit 32: FBN's initial goal is to leverage data and ML to help farmers make informed decisions
Cycle of data aggregation and use



Source: Company data, Goldman Sachs Global Investment Research

Exhibit 33: With the ultimate goal of replacing procurement, sales, and finance incumbents



Source: Company data, Goldman Sachs Global Investment Research

Financial Services

$34-$43bn annual cost savings & new revenue opportunities by 2025

Machine learning and AI have broad applications in the financial services sector, as the presence of robust, rich data sets informing investment decisions and credit risk profiles illustrate an environment conducive to usage of algorithms in boosting data efficiencies. The ability for machine learning technology to leverage pattern recognition in a fraction of the time of a human-driven equivalent provides opportunities for the procurement and analysis of unique data to more accurately inform investment decisions. Furthermore, the availability of vast, comprehensive proprietary financial data within commercial banks provides opportunity for AI/ML usage in reducing costs for the general banking sector. **Conservatively, we believe machine learning and AI could enable access to roughly $34-$43 billion per year in cost savings and new revenue opportunities by 2025, with further upside as these technologies enable faster and more complex data leveraging and execution.**

What is the opportunity?

Maximizing investment potential. Money managing firms with relative technological leverage (i.e. quant hedge funds) are best positioned to take advantage of competitive profit opportunities using machine learning techniques, in our view. Recent developments connecting deep learning algorithms to application accelerators have improved speed and efficiency of recognizing trends across data and even image sets, providing a clear path forward for firms looking to obtain a competitive edge in information and execution leverage.

On the data side, we believe that AI/ML could provide significant advantage in analysis for informing investment decisions, creating opportunities for both cost reduction and breaking into new profit pools. On the execution side, the more than 1.7 trillion U.S. equity shares traded in 2015 highlights plenty of opportunity for trading firms to take advantage of the miniscule latency windows where the latest price of a security exists on raw exchange feeds but not consolidated market systems, an area we believe AI/ML can make a meaningful difference. We believe that by leveraging cost-effective hardware accelerators and pattern-recognition capabilities, AI/ML could have a significant impact on data quality as well as analysis, procurement and execution speed, leading to $19-$28bn benefits per year from better informed investment decisions and quicker reaction to market events by 2025.

Reducing credit risk. For traditional lending institutions, we believe AI/ML has the potential to meaningfully reduce credit risk through these advantages, identifying accounts at risk and executing credit line reductions/eliminations that could reduce balance sheet charge-offs and loan loss reserves for these institutions. Even in an environment with relatively low charge-off rates, growing consumer credit outstanding has resulted in ~$60bn in consumer credit related charge offs per year according to the Federal Reserve Board, that can be reduced by 19% by 2025 using AI/ML, in our view.

Reducing compliance and regulatory costs. For financial services firms such as community banks and large investment banks, we estimate compliance-related employee costs alone to be upwards of $18bn per year. While many firms have seen a 50% or more increase in compliance costs in the past few years, we believe AI/ML has the potential to meaningfully reduce this cost burden on the industry. Including credit risk reduction, we believe AI/ML can provide a ~$15bn per year cost reduction opportunity in the financial services sector by 2025.

What are the pain points?

Today, firms face the resource allocation dilemma of balancing employee compensation and capital investment in evolving technologies, with the purpose of expanding usage of "good data" to generate returns on capital and suppressing costs. Below are three pain points we believe hinder firms' abilities to effectively leverage data:

Latency Arbitrage refers to a trading strategy where firms attempt to trade based on data more quickly than the market systems can respond to that data.

Execution speed. A major pain point for asset managers, especially high frequency traders (HFT) who trade on technical market movement, is staying competitive in a highly liquid and fast-paced market where milliseconds determine wide variance in return potential. For example, the median length of latency arbitrage windows was almost a full second in 2014 and has reduced significantly since, highlighting the steepening ramp for money managers with exposure to arbitrage or HFT strategies.

Data access. On the fundamentals side, we view a wide variety of useful data as unreliable or unattainable due to measurement limitations, geopolitical restrictions, and analysis cost constraints. High technological barriers have prevented asset managers looking to gain competitive advantages from accessing novel, timelier, and more accurate data.

Cost duality. In our view, labor costs for data scrubbing, analysis, and execution have played a significant role in keeping asset management operating margins below 40% in the last decade. Moreover, start-up non-recurring engineering (NRE) costs of programmatic acceleration hardware have historically provided cost barriers to leveraging technology to increase competitiveness. Growing availability and flexibility of lower cost options such as Field Programmable Gate Arrays (FPGAs) provide more accessible avenues for AI/ML processes to be leveraged.

What is the current way of doing business?

Human capital drives cost, risk-management structures. For many large asset managers today, revenue-generating employee costs make up between 1/3 and 1/2 of revenue generated as employees are responsible for sifting through robust data sets, management commentary, and research views to make informed investment decisions that benefit clients. To the same effect, loan officers at traditional lending institutions are often responsible for approving and monitoring credit revolvers and term-loans for potential delinquency, with the general responsibility of minimizing firm loan losses. At investment and community banks, too, the evolving regulatory environment has increased capital spend in compliance efforts that require human capital.

Market reliant on scheduled primary source data releases to gauge comprehensive ROIC. With low barriers to big data access and low latency diffusion of relevant one-off market events through online channels, investors allocate large amounts of labor and capital towards efficiently scrubbing data sets, gaining a proprietary edge, and reacting quickly to rapidly evolving circumstances. Regardless of these proprietary advantages, however, money managers are ultimately reliant on primary source data releases (e.g. weekly EIA oil inventory data, company earnings reports) to gauge forecasting ability, subsequent market moves, and resulting ROIC.

How does AI/ML help?

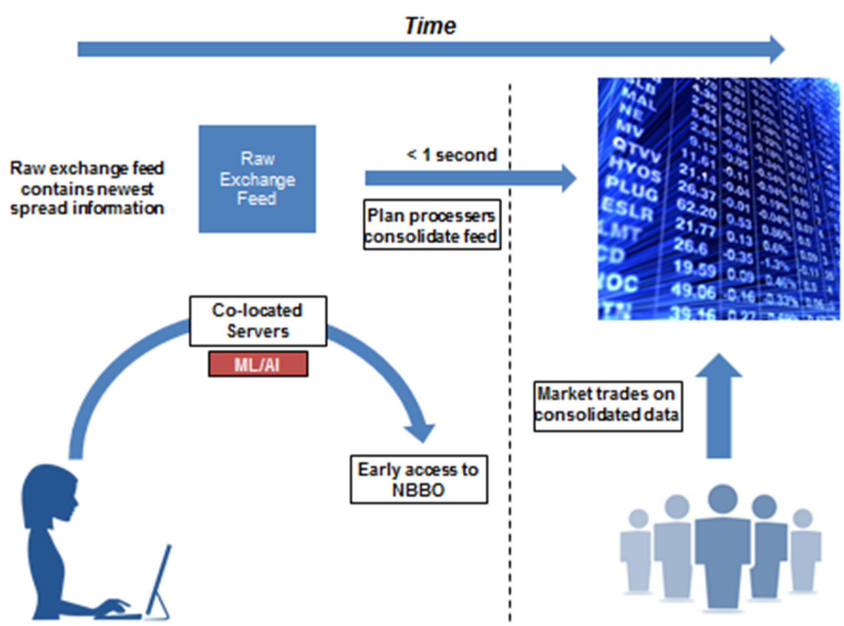
Machine Learning's applications can quickly monitor and process robust data sets in seeking to analyze and/or execute on specific end goals, which particularly suits high-frequency trading firms, traditional asset managers and traditional lending institutions.

Execution speed. Asset managers with a HFT focus have faced increasing pressure from competitors as evolving adoption of technologies has reduced reaction time to technical and one-off fundamental market catalysts. Latency arbitrage is one practice that funds are using to gain access to trading information mere fractions of a second before the market; mitigated through increasing optionality of hardware accelerators such as ASICs and FPGAs.

National Best Bid and Offer (NBBO) refers to the lowest available ask price and the highest available bid price. Brokers are required to execute the best available price when the customer buys or sells securities.

Firms are able to reduce latency in two distinct ways. First, they are able to co-locate trading servers at the exchange, reducing physical distance and enabling quicker acquisition of relevant trade-data. Second, these firms are able to source data from raw exchange feeds and retrieve the national best bid/offer (NBBO) prices faster than traditional data consolidation processes (Exhibit 37). Firms can have a clear edge by receiving data slightly before the market, and we believe machine learning algorithms have the potential to more quickly and accurately identify and execute on the price spread before the latency period is exhausted.

Exhibit 34: Latency arbitrage provides early NBBO Access
AI/ML facilitates data capture and execution



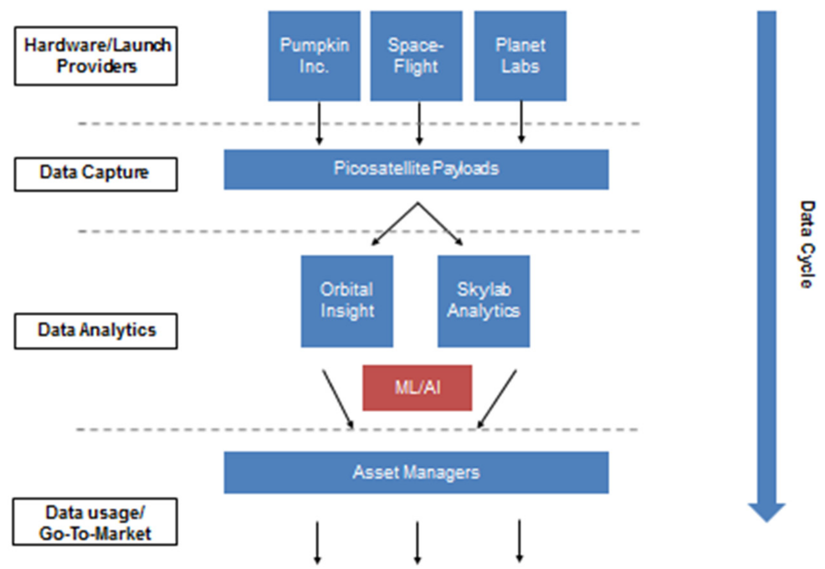
Source: Goldman Sachs Global Investment Research

Data access. As technological evolution facilitates access to big data for traditional asset managers, companies are increasingly trying to find competitive advantages in the industry. Data analytics firms are entering the fold to capture untapped opportunities. For example, some firms are utilizing data from satellites that capture images of areas providing information about equities, commodity prices, and even full-scale economies. For firms like Cargometrics and Orbital Insight, these images include shipping patterns to inform commodities prices to store parking lots, and informing customer growth rates at retailers, respectively. A few companies are building their own rockets and booking future launches for small satellite payloads, whereas companies like SpaceFlight guarantee launches by working with launch vehicle providers around the world.

Data analytics firms leveraging machine learning/AI are utilizing image recognition abilities of algorithms, such as convolutional neural networks (CNNs), to scrub image data for specific characteristics in specific regions of the world. In this way, they are able to more quickly and accurately tailor data from sensitive, remote, and dense regions and package it to inform specific market trends. The VC firm Deep Knowledge Ventures exemplifies industry commitment to big data, most notably through appointing a data analysis algorithm, known as VITAL, to its Board of Directors in 2014.

The picosatellite data cycle

AI/ML creating advantage over traditional data collection



Source: Goldman Sachs Global Investment Research

Credit Risk Reduction. Charge-offs hinder commercial bank balance sheets and cash flows, which we estimate to be ~$60bn per year related to consumer credit. Based on a report from Khandani et al. (*"Consumer credit-risk models via machine learning algorithms,"* MIT, 6/10/2010), we believe AI/ML has the potential to promptly identify risk in revolving lines of credit (RLOCs) and execute limit reductions/eliminations with accounts that data suggests could go delinquent.

The research suggests that their machine learning model was able to forecast credit delinquencies in RLOCs with a linear regression R^2 of 85%, highlighting ML applications for scrubbing and executing on credit data. We further believe that ML could contribute similarly to non-revolving consumer loans by assisting loan officers in determining creditworthiness beyond the use of typical metrics. In terms of fraud detection, too, companies such as AIG and Stripe, a private payments company, are using machine learning advances to better inform and determine patterns in fraudulent activity claims and transactions.

Reducing Compliance Costs. On the compliance side, small community banks and large investment banks alike are boosting spend to ensure vigilance of new regulations facing the industry. According to J.P. Morgan's most recent annual report, the company increased compliance spend by 50% to $9 billion between 2011 and 2015. To the same tune, Citigroup indicated in 2014 that their compliance employee headcount grew to 30,000, representing over 12% of their employee base.

We believe that AI/ML could have a meaningful impact in reducing the employee overhead required to do certain tasks. For Digital Reasoning, a private analytics firm based in Nashville, machine learning techniques are developed to provide proactive compliance analytics, completing tasks such as sifting through employee emails for potentially non-compliant content and detecting violations such as market manipulation, unauthorized trading or wall cross violation.

Data Access & ROI. To illustrate the potential effects of gaining access to comprehensive, proprietary data using AI/ML on ROI potential, we conducted an analysis with oil futures investments and isolated front-end oil futures contract prices from 2011-2016. Using contract price data, we found a 14% increase in volatility in the oil futures market on the day of EIA oil storage data releases (Wednesday, weekly).

Given the value and insight gained from procuring ML scrubbed, hi-res imagery data of oil containers, rigs, shipping movements, and production facilities without geographical or geopolitical constraints, it is our view that the financial services industry has the opportunity to take advantage of data-driven market events using machine learning. The volatility around data-releases in the oil futures market is an example of how better data can be used to inform investment decisions and provide the potential for better returns.

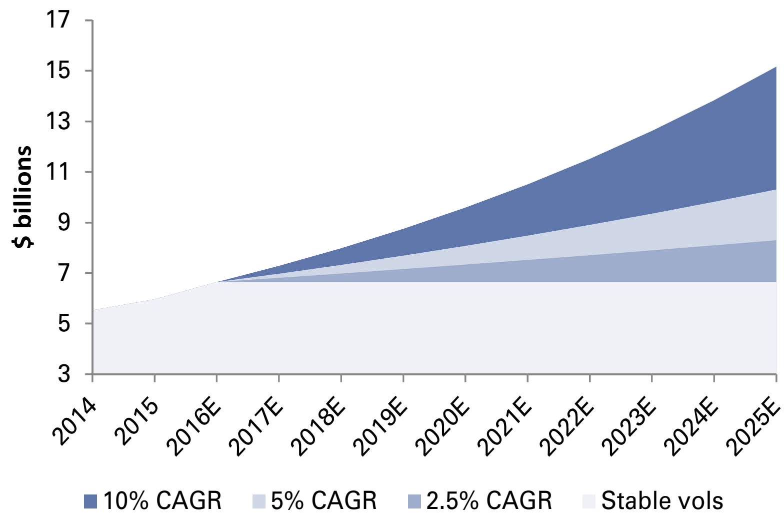
Quantifying the opportunity

We estimate that AI/ML has the potential to facilitate roughly $34-43bn per year in cost reduction and new revenue opportunities for the financial services industry by 2025, with more upside potential as related technologies evolve in complexity and sophistication. We quantify AI/ML contributing $6.5bn-$15bn per year in untapped latency arbitrage opportunities, $13bn per year in asset manager operating cost reduction from more efficient data access, ~$2bn per year in compliance cost reduction, and ~$13bn per year in charge-off reductions for traditional lending institutions.

Latency Arbitrage. In order to quantify latency arbitrage potential in the U.S. equity market, we leverage academic research by Elaine Wah (*"How Prevalent and Profitable are Latency Arbitrage Opportunities on U.S. stock exchanges?," University of Michigan, 2/8/2016*). According to the researcher, total latency arbitrage profitability in 495 S&P 500 stocks was $3.03bn in 2014, with each stock having approximately 69 arbitrage opportunities per day.

Using this analysis, we find that such a profit equates to over $^{3}/_{10}$ of a penny per share traded, and when extrapolated to total estimated U.S. equity volumes in 2016 yields a profit pool of $6.5bn. Assuming equity volume growth remains constant vs. 2014-2016 levels (10% CAGR), this implies a $15bn annual profit pool by 2025.

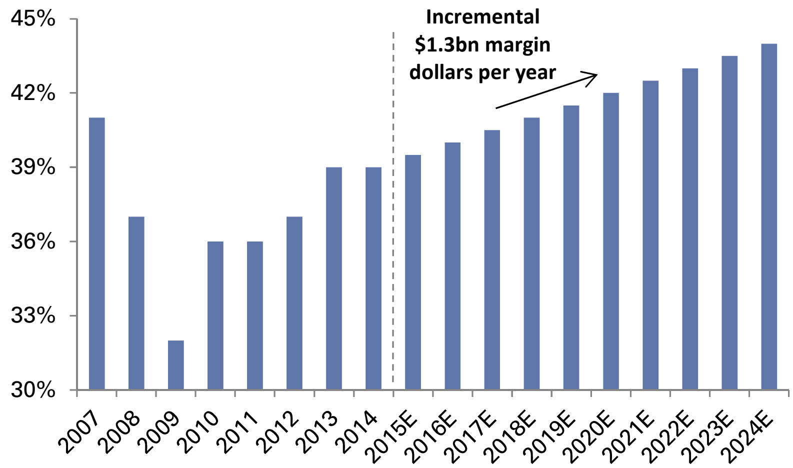
Exhibit 35: Latency Arbitrage Profit Pools
$6.5bn-$15bn by 2025 depending on CAGR of equity volumes



Source: NYX Data, Goldman Sachs Global Investment Research

Asset management cost reductions. The Boston Consulting Group has noted that asset manager operating margins as a % of net revenues have stayed flat at 39% in recent years after reaching 41% in 2007, and profits reached $102bn in 2014. We believe that introduction of AI/ML in data access and analysis, as highlighted in our picosatellites case study, will slowly allow hedge funds and other asset managers to reduce labor needs at a faster rate than the growth in data procurement costs. All else being equal, we expect a 5% reduction in operating costs, or an incremental $1.3bn per year, for the asset management industry in the next decade. This implies a $13bn annual cost reduction by 2025 for the industry, and may prove conservative in its assumption that hedge fund profits remain roughly flat in upcoming years. Given the level at which asset managers leverage human capital, we believe that AI/ML could bring operating margins higher than we saw in 2007.

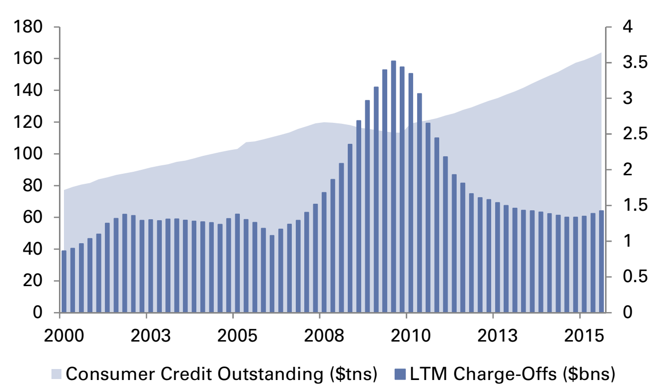
Exhibit 36: Asset Manager operating margins as % of net revenue have stagnated
We expect marginal increases through 2025 with AI/ML



Source: Boston Consulting Group (Global Asset Management 2015, July 2015), Goldman Sachs Global Investment Research

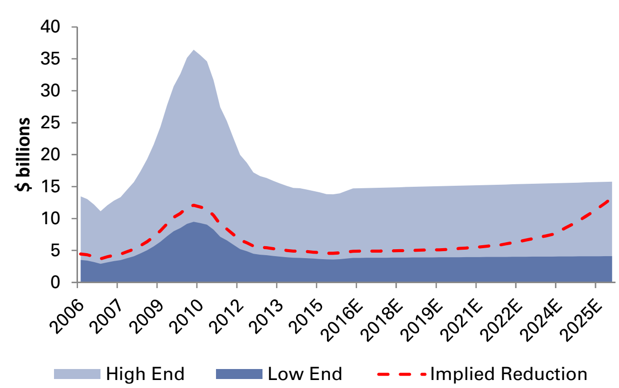
Traditional lenders and risk reduction. In the work by Khandani et al. on ML and consumer credit risk cited earlier, the researchers indicated their machine learning model for revolving consumer credit delinquency implied charge-off cost-savings between 6% and 23%. We remain closer to the conservative end of this range initially (implied cost-savings of ~8%) with our addition of non-revolving loans to ML applications, given that machine learning would, in our view, only be useful on the front end of loan approval rather than throughout the payment schedule like in revolving credit situations. By 2025, however, our implied cost-savings increases to 19% as technology grows more sophisticated in outer years. Assuming equal delinquency likelihood in each category, ¼ of charge-offs come from revolving credit and the other ¾ from non-revolving (NR) credit agreements. Based on these assumptions, we expect AI/ML to contribute ~$13bn per year in cost savings for traditional lending institutions by 2025.

Exhibit 37: Consumer credit continues to grow
And commercial bank charge-offs have stabilized at ~$60bn



Source: Federal Reserve Board, Goldman Sachs Global Investment Research

Exhibit 38: Machine Learning could reduce loan losses
Our estimates imply ~$13bn per year cost reduction by 2025



Source: Federal Reserve Board, Goldman Sachs Global Investment Research

Reducing Compliance Costs. We estimate that community banks and large investment banks pay a total of ~$18bn in compliance-related employee costs per year. According to a Federal Reserve study on Community Banking in the 21st Century, community banks (banks with <$10bn in assets) paid over $3bn in 2015, while we estimate the largest 10 investment banks paid just under $14bn. We assume an average salary per compliance employee at $69,000 per year and believe that AI/ML could provide a 10% reduction in compliance employee costs, as portions of bank compliance efforts become driven by machine learning developments. Based on these assumptions, we believe AI/ML can contribute ~$2bn per year in compliance cost reduction for banking firms by 2025.

Who could be disrupted?

Companies with capital constraints and traditional asset management practices could be disrupted as adapting firms invest more in competitive AI/ML trading hardware and new pools of proprietary data. As these firms reduce latency in closing market inefficiency gaps, there could be less room for firms that rely solely on human capital for research and for trades on technical market movement/one-off fundamental market catalysts.

Companies with burdensome credit approval processes or which rely on a small number of creditworthiness metrics when approving loans (i.e. credit scoring only) could potentially suffer as competitors begin to use AI/ML. As machine learning algorithms reduce/eliminate risky credit lines at cutting edge firms, those customers could increasingly attempt to receive loans from traditional institutions without ML safeguards, disproportionately putting these firms at higher risk for increased delinquencies.

Healthcare

$54bn annual cost savings by 2025

defined datasets, the need for monitoring over time, and the wide variability of outcomes offer the potential for disproportionately high returns on the technology implemented in areas like drug discovery, test analysis, treatment optimization, and patient monitoring. With the integration of machine learning and AI, the opportunity exists to significantly **de-risk the drug discovery and development process, removing $26bn per year in costs, while also driving efficiencies in healthcare information worth more than $28bn per year globally.**

What is the opportunity?

Virtual screening is a computer-based method used in drug discovery to identify structures within a vast library of small molecules that are most likely to have a particular effect on the drug target

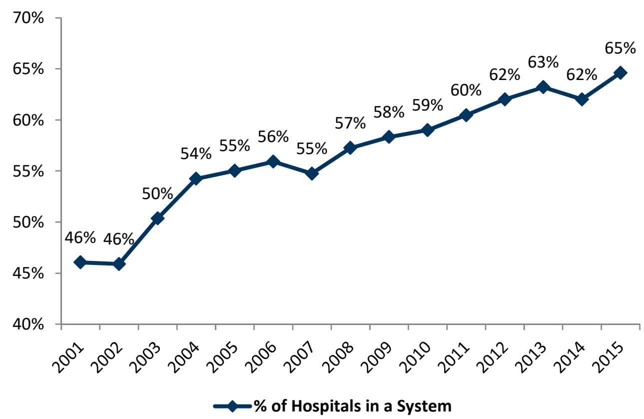
Drug discovery and development. The potential efficiency gains from incorporating machine learning processes throughout development could not only accelerate the time horizon but also improve returns on R&D spend by increasing the probability of success (POS) for medicines reaching late-stage trials. According to David Grainger, Partner at Medicxi Ventures, avoiding the False Discovery Rate, a mostly statistical driven phenomenon according to him, could de-risk late stage trials by half. Further, the current method of virtual screening in early stage drug discovery known as high throughput screening is highly vulnerable to this type of statistical error. Halving the risk of expensive Phase 3 trials could generate billions in savings and meaningfully impact returns on the more than $90bn in R&D spend across the largest pharmaceutical companies, freeing up resources to focus on finding higher potential opportunities.

While substantial costs associated with late-stage trials are often focused in clinical trial design elements, we believe meaningful efficiency gains can also be realized throughout later stages with AI/ML implementation to optimize decisions around selection criteria, size, and length of study.

Doctor/hospital efficiency. Driven partly by regulation and fragmentation, the healthcare system in the US has historically been slow to adopt new technologies. Apart from the systematic challenges, the time between new discoveries and when doctors and clinics put new medicines or treatments to use is often long and inconsistent.

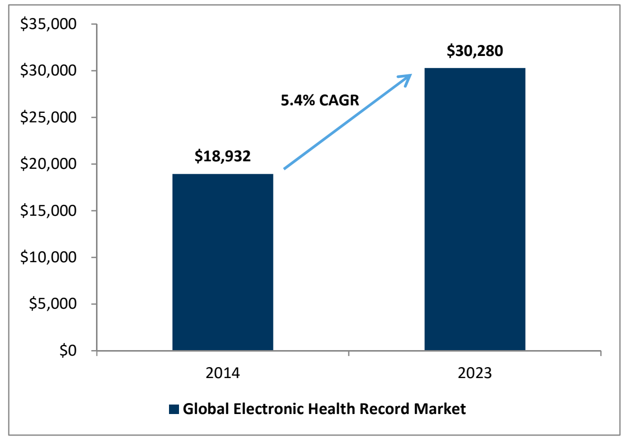
Recent mandates from the US government as part of the American Recovery and Reinvestment Act have driven growth in spaces like electronic health records, a global market expected to reach ~$30bn by 2023, according to Transparency Market Research. The aggregation of data, improving technology to capture it, and secular decline of standalone hospitals has created an opportunity to leverage data at a scale not attainable historically. This in turn is enabling machine learning algorithms and AI capabilities to demonstrate early traction improving the speed, cost, and accuracy in various areas of healthcare.

Exhibit 39: Provider consolidation
% of community hospitals in a health system



Source: American Hospital Association

Exhibit 40: Electronic health records market expanding globally
$ millions



Source: Transparency Market Research

Google's DeepMind division headquartered in London is collaborating with the UK's National Health Service (NHS) to build an app aimed at monitoring patients with kidney disease as well as a platform formerly known as 'Patient Rescue' that seeks to support diagnostic decisions. A key input for any AI/ML system is immense amounts of data, so DeepMind and the NHS entered a data-sharing agreement providing DeepMind with a continuous stream of new data and historical records it is leveraging to train its algorithms. This real-time analysis of clinical data is only possible with vast amounts of data, though the insights provided by DeepMind's effectively unlimited access to patient data could deliver learnings well beyond the scope of kidney disease.

What are the pain points?

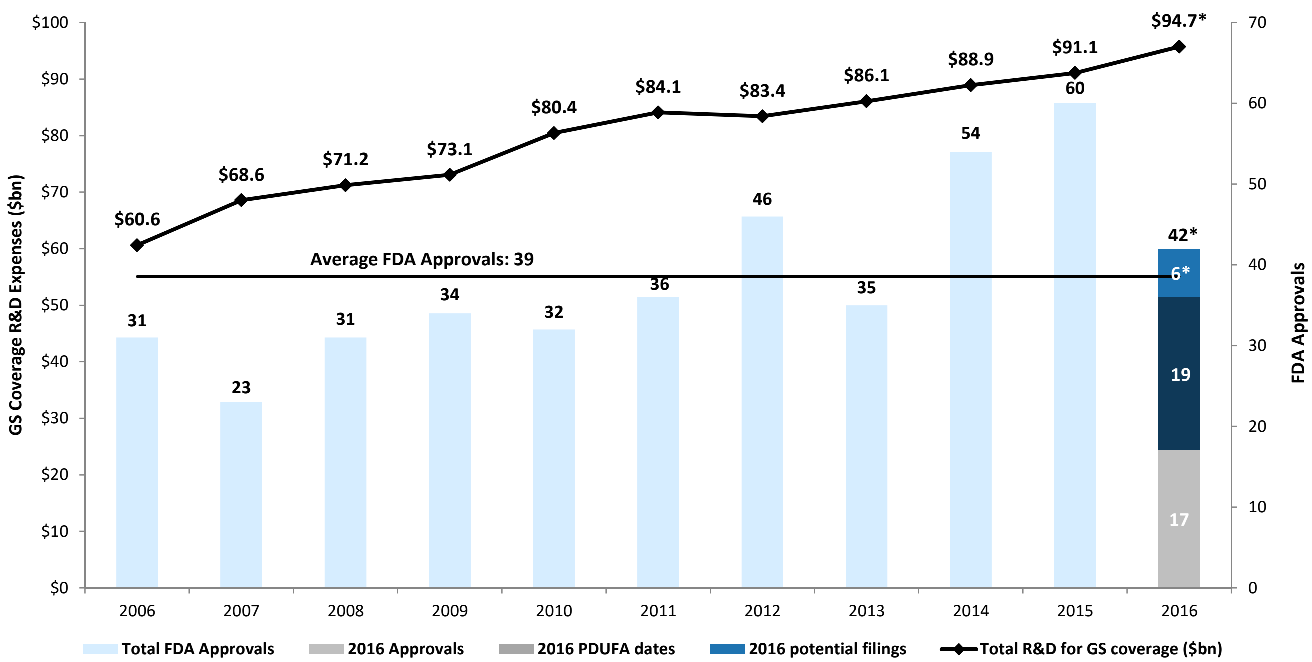
Drug discovery and development. A significant pain point within healthcare is the time and cost of drug discovery and development. It takes approximately 97 months on average for new therapies to progress from discovery to FDA approval, according to the Tufts Center for the Study of Drug Development. While a focus on specialty can aid the time horizon, costs have continued to increase steadily as well. Deloitte found that across a cohort of 12 major pharmaceutical companies the cost to develop an approved asset has increased 33% since 2010 to roughly $1.6bn per year.

R&D returns. R&D productivity in biopharma remains a debated topic. While the cost of developing successful drugs has increased, the revenue environment has also been unsupportive of returns due to reimbursement headwinds, lower patient volumes, and competition. While we expect returns to improve from 2010-2020 vs. 2000-2010, the change is marginal. Further, one of the most significant headwinds to returns remains failed assets, particularly those reaching later stages, which we estimate account for more than $40bn in annual costs.

Doctor/hospital efficiency. A challenge unique to healthcare remains the significant lag between when new drugs and treatments are approved versus when doctors begin implementing with patients. As a result, many machine learning and AI experts working in healthcare continue to encourage major providers to integrate modern machine learning tools into their workflows which can fully utilize the vast stores of medical data being collected and published today.

Opportunities exist for machine learning and AI to decrease the time between discovery and application and also optimize treatment. For example, a 2009 study from the Radiological Society of North America on hepatobiliary (liver, gall bladder) radiology found that 23% of second opinions had a change in diagnosis, a problem machine learning companies focused on medical imaging have an opportunity to solve. Further, companies like Deep Genomics that use machine learning to identify diseases at the genome level are positioning providers to deliver more targeted and effective treatment.

Exhibit 41: Large-cap biopharma 10-year R&D spend vs. FDA approvals
$ billions



Source: FDA, Company data, Goldman Sachs Global Investment Research.
Note: Total R&D spend includes large cap US biotech (ALXN, AMGN, BIIB, CELG, GILD, REGN and VRTX), large cap US pharma (ABBV post-ABBV split, BMY, JNJ, LLY, MRK and PFE) and large cap EU pharma (AZN, BAYGn, GSK, NOVN, NOVOb, ROG and SASY).

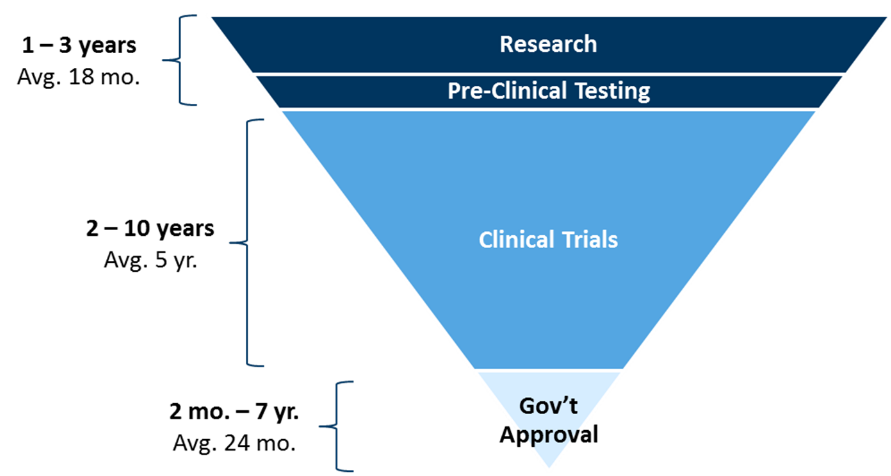
What is the current way of doing business?

The current drug discovery and development business is an extensive process of research, testing, and approval that can last more than 10 years. Time to market analysis from the Tufts Center for the Study of Drug Development reports it takes 96.8 months on average for a drug to advance from Phase 1 to FDA approval. Discovery of new treatments is a unique challenge not only because of the length of time required but also because of the low POS throughout the various stages of development.

Drug discovery initially begins with identifying a target. Once a target has been identified, high throughput screening (HTS) is often used for “hit finding”. High throughput screening (HTS) is an automated, expensive process carried out by robots that tries to identify these “hits” by conducting millions of tests to see which compounds show potential with the target. The hits then transition to lead generation where they are optimized to find lead compounds, which are then optimized more extensively before progressing to pre-clinical drug development. This entire process can last 1-3 years before a drug reaches Phase 1, at which time it is understood to have only a 20% probability of success.

- Phase I: Emphasis on safety; healthy volunteers (20% POS).
- Phase II: Focuses on effectiveness; volunteers with certain diseases or conditions (40% POS).
- Phase III: Further information gathered on safety and effectiveness across different populations, dosages, and combinations. Ranges from several hundred to thousands of volunteers (60% POS).

Exhibit 42: Drug discovery and development timeline



Source: Princess Margaret Cancer Foundation, Goldman Sachs Global Investment Research

How does AI/ML help?

The advantages and use cases of machine learning and AI within the healthcare industry are wide ranging. Not only are decisions driven by data, rather than human understanding or intuition, but the decisions and/or predictions are able to consider a combination of factors beyond human capacity. Deep learning in particular shows unique potential as it can exploit knowledge learned across different tasks in order to improve performance on others tasks.

Reduce failed discovery and increase POS. Significant capital is invested with substantial opportunity cost in exploring treatments that are understood to have roughly 20% probability of success (POS) if they reach Phase 1 trials. As a result, AI/ML has been applied, almost entirely within academics to date, in an effort to develop efficient and accurate virtual screening methodologies to replace costly and time consuming high throughput screening processes.

Google and Stanford researchers recently leveraged deep learning in an effort to develop virtual screening techniques to replace or augment the traditional high throughput screening (HTS) process and improve the speed and success rate in screening. By applying deep learning the researchers were able to facilitate information sharing across multiple experiments across multiple targets.

> ***"Our experiments have shown that deep neural networks outperformed all other methods…***
> *In particular, deep nets surpassed existing commercial solutions by a large*

margin. On many targets it achieves nearly perfect prediction quality which qualifies it for usage as a virtual screening device… In summary, deep learning provides the opportunity to establish virtual screening as a standard step in drug design pipelines." (Massively Multitask Networks for Drug Discovery, 2/6/2015)

Merck hosted a Kaggle challenge in 2012, also aimed at identifying statistical techniques for virtual screening and is beginning to test applications of deep learning and AI, specifically through partnering with AI drug discovery startup, Atomwise. Atomwise recently leveraged AI technology trained to analyze drugs as a chemist to understand how safe, existing drugs could be repurposed to treat Ebola. The analysis, which assessed ~7,000 existing drugs, was performed in less than one day. Historically, this analysis would have taken months or years to perform, according to the company.

Improve doctor/hospital efficiency. Early successes in applying machine learning have been seen with improving diagnoses (Enlitic, DeepMind Health), analyzing radiology results (Zebra Medical Vision, Bay Labs), genomic medicine (Deep Genomics), and even the use of AI in treating depression, anxiety, and PTSD (Ginger.io). As health data becomes more accessible as a result of both the digitization of health records and aggregation of data, significant opportunity exists for AI/ML to not only remove costs associated with procedural tasks but also improve care via algorithms that let historically disparate data sets communicate. Ultimately, the capabilities of AI/ML to consider factors and combinations beyond human capacity will allow providers to diagnose and treat with greater efficiency.

Broad Institute of MIT and Harvard: Using AI/ML in genomics and the fight against cancer

Broad Institute of MIT and Harvard, a not-for-profit biomedical and genomic research center in Cambridge, MA, sits at the intersection of academia and industry. Through its affiliation with Harvard and MIT the center fosters collaborative research, cutting across different domains, with the goal of publishing its findings or licensing them to biotech or pharmaceutical companies. Ultimately, the center aims to produce things that people can use to advance their scientific agendas, according to **Broad Institute Chief Data Officer, Cardiologist, and Google Ventures partner Anthony Philippakis**.

We highlight the following key takeaways from our conversations:

Opportunities, challenges. The Broad Institute is currently running man vs. machine tests in certain areas of its genomic research. To date, matching algorithms have not been wrong and have also substantially reduced the workload on a group of humans. Still, humans are likely to remain involved where data or results might involve ethical or political issues.

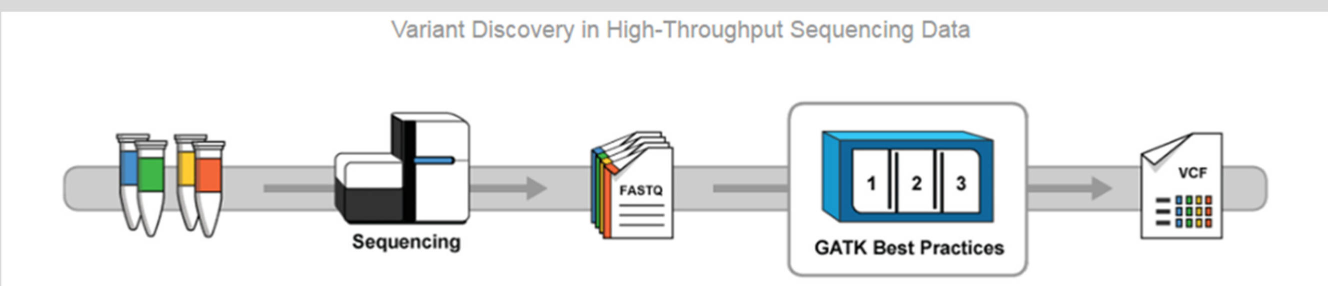
While there is no shortage of use-cases for AI/ML to save lives or make a discovery, finding the right business model to support the end-goal is the main challenge, according to Mr. Philippakis. Currently, doctors are being asked to make life-changing decisions every day based on remembered facts and acronyms to determine treatment paths, an area of significant opportunity for AI/ML within clinical care. However, significant impediments exist as reimbursement for decision support tools and incentives for opening up the health data necessary for this level of AI/ML integration have not caught up.

Technology stack. Broad is a substantial user of Google Cloud with a large amount of its platform open-sourced as the institute's teams aim to build within a Spark Framework. Though the technologies are still largely in the transition phase, the progression to cloud vendors has been rapid as they are the only ones able to match the levels of growth. In addition, many cloud vendors are building out dedicated teams for genomics, though the evolution remains at the beginning of the learning curve on "ML-as-a-Service" products, according to Mr. Philippakis.

Data, data, data. Broad institute generates roughly a petabyte of data per month with a doubling time of every 8 months. As a result, the institute is partnering with companies like Cycle Computing to move away from legacy data handling methods. With structured data comes meaningful follow-on opportunities and the institute is also pushing along other data and analytics standards like its Genome Analysis Tool-Kit (GATK) which offers a wide variety of tools focused on variant discovery and genotyping.

Ideally the scientists at Broad would analyze genomic data alongside EHR (electronic health records) to understand the relationship between specific cell lines and cancer, but limited incentives exist in the EHR world today to open-up and share, as most incentives are set up around risk and data disclosure. That said, Mr. Philippakis sees Broad operating in a world of mostly open genomic data in the future unlike other verticals that have aimed to hoard data.

Exhibit 43: Broad Institute's Genome Analysis Toolkit
Focus on variant discovery and genotyping



Source: Geraldine Van der Auwera / Broad Institute of MIT and Harvard

Quantifying the opportunity

Cost of failed discovery. We analyze the impact of halving the risk associated with drug development discovery through the implementation of machine learning and AI under the following assumptions:

- The average annual cost of development for an approved asset is $1.6bn, inclusive of costs associated with failed assets (Deloitte).
- The $30bn in annual costs from failed assets can be distributed evenly across the number of approved assets reported by the cohort analyzed, or 43 (Deloitte).

The FDA reported 60 approvals in 2015 which would imply, based on the cost of failed assets per approved asset (~$698mn in 2015), nearly $42bn was allocated to failed assets. We assume machine learning and AI could halve the risk of the development process, producing roughly **$26bn in annual savings within the pharmaceutical industry globally** by 2025.

Exhibit 44: AI/ML could eliminate ~$26bn in development costs
$ millions

ML/AI drug discovery/dev. cost opportunity	
Cohort annual failed assets cost	$37,175
Approved assets	43
Failed asset costs per approval	$865
FDA approvals	60
Total cost of failed assets	$51,872
ML/AI cost savings opportunity ($mn)	**$25,936**
Annual approved asset cost ($mn)	$2,567
Failed asset cost	$865
% of total cost	**34%**

Source: Deloitte, FDA, Goldman Sachs Global Investment Research.

Accelerating gains from a shift to electronic health records. In the US alone, health information technicians represent ~$7bn in annual payroll today. Driven in part by an ageing population and government mandates to transition to digital, the job outlook for health information technicians is expected to experience above average growth from 2014-2024, according to the BLS, expanding 15% vs. 7% across all other occupations. However, given the susceptibility to automation and substitution via software of many tasks within the occupation, we believe machine learning and AI could potentially displace nearly all of these jobs.

Health information technicians ensure the quality, accuracy, accessibility, and security of patients' health data for reimbursement and/or research, while also leveraging technology to analyze patient data in order to improve care and control costs, according to BLS. A proliferation of AI/ML within the healthcare industry would likely have serious implications for occupations such as this, and we estimate based on per capita healthcare spend and share of global spend that AI/ML could **remove more than $28bn in annual costs globally** by 2025.

Exhibit 45: AI/ML could displace nearly all health information tech (HIT) positions
$ millions

Health Information Technicians (HIT)	
US Median annual pay	$51,636
Number of jobs in US	218,776
US Annual cost ($mn)	**$11,297**
US healthcare spend ($mn)	$2,998,469
Global healthcare spend ($mn)	$7,536,116
US share	40%
Global HIT cost ($mn)	**$28,392**

Source: Population Reference Bureau, World Bank, BLS, Goldman Sachs Global Investment Research.

Who could be disrupted?

We believe machine learning and AI have the potential to dramatically change the big pharma landscape, and healthcare systems more broadly, based on cost savings and POS improvements throughout drug discovery and development as well as efficiency gains across providers and facilities. We would expect, over the long-term, a proliferation of machine learning and AI technologies to increase competition within drug development as time horizons shorten and losses on failed assets decline.

Further, efficiency gains and automation could prove disruptive for medical professions and companies that over-index to interpreting results and making diagnoses versus the actual delivery of care or performance of surgery, such as radiologists, specialists offering second opinions, and administrative or support staff. We believe that most of this disruption is longer-term, as we are still in early development stages for many of these technologies and the cost to early adopters is potentially prohibitive, relative to other improvement mechanisms.

Challenges to adoption

While the opportunities for AI/ML within healthcare span across many sub-sectors, barriers to adoption remain.

- **Cost**. Costs for necessary tools and capabilities that serve as prerequisites for AI/ML could prove prohibitive, particularly within healthcare where cost of care remains a focus. Investments to ensure ML algorithms are leveraging good data require meaningful capital and know-how and the compute power alone will prove costly.
- **Interpretability.** Algorithms combing across multiple data sets can produce somewhat of a black box. Industries like healthcare that have historically been more heavily regulated could push back on the advancement of AI/ML applications as a result.
- **Talent.** Barriers to adoption could also stem from a concentration of talent capable of applying AI/ML and interpreting results. In 2013, Google paid more than $400mn to acqui-hire DeepMind Technologies; according to press reports, a team of roughly a dozen. This consolidation of talent and the resulting cost could prove prohibitive.
- **Data**. While government mandates have helped to digitize electronic records in the US, challenges remain in transitioning paper-intensive systems to fully electronic. Further, while many have reached the “meaningful use” threshold, fragmentation and lack of accessibility to important patient data could impede progress.

Retail

$54bn annual cost savings & $41bn annual new revenue by 2025

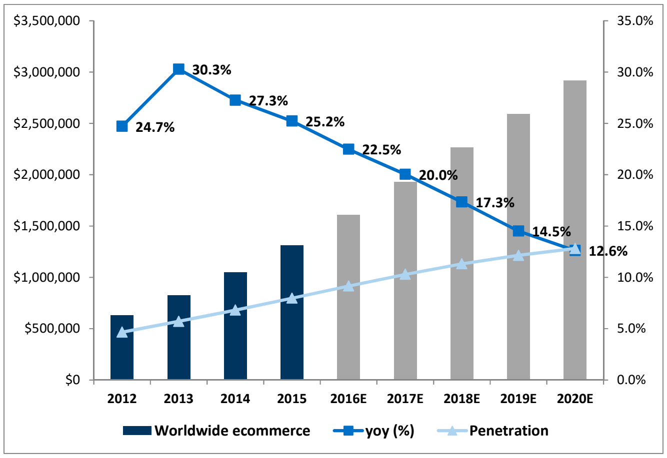
category, the advent of ecommerce has also generated massive amounts of customer data for retailers. Still, the most important question remains; how do companies leverage the data they have on hand to serve customers better and make more money? Early signs of success were seen with the proliferation of ad tech which allowed retailers to more efficiently and effectively target customers across the web. Today, retailers are leveraging historically disparate data sets to optimize not only advertising but also inventory management, demand prediction, customer management, and trend extrapolation. **We see the opportunity for AI/ML to further these efforts by predicting demand and driving labor efficiencies worth $54bn annually while also optimizing pricing and generating annual sales lift in discretionary categories like apparel and footwear of $41bn** globally by 2025.

What is the opportunity?

Retail as a sector is navigating significant secular trends as millennials move into their prime purchasing years and consumer buying habits shift online. While retailers have navigated these challenges with varying degrees of success to date, AI/ML presents an opportunity for omnichannel and pure-play ecommerce retailers to pull insights from the massive amounts of customer and product data accumulated as purchases shift online and technologies improve. Within our research we have identified key areas of opportunity for AI/ML that span the retail value chain.

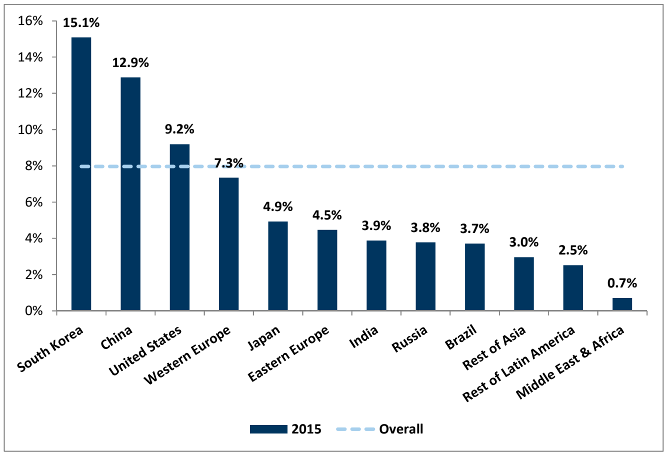
While recommendation engines are not a new phenomenon within ecommerce, traditional techniques face certain limitations that we believe AI/ML processes could surpass to provide deeper, more accurate insights from both sales and content data. In addition, natural language processing (NLP) AI systems are enabling more intuitive and relevant search as well as conversational commerce. Further, the integration of AI/ML into both the earlier stages of wholesale and retail buying and the later stages of selling could drive greater labor and inventory efficiencies as a result of more precise demand prediction and improve sales with optimized pricing.

Exhibit 46: Global ecommerce
17% 2015-2020E CAGR, $ millions



Source: Euromonitor, Goldman Sachs Global Investment Research.

Exhibit 47: Ecommerce penetration by country
8% overall penetration in 2015



Source: Euromonitor, iResearch, METI, Goldman Sachs Global Investment Research.

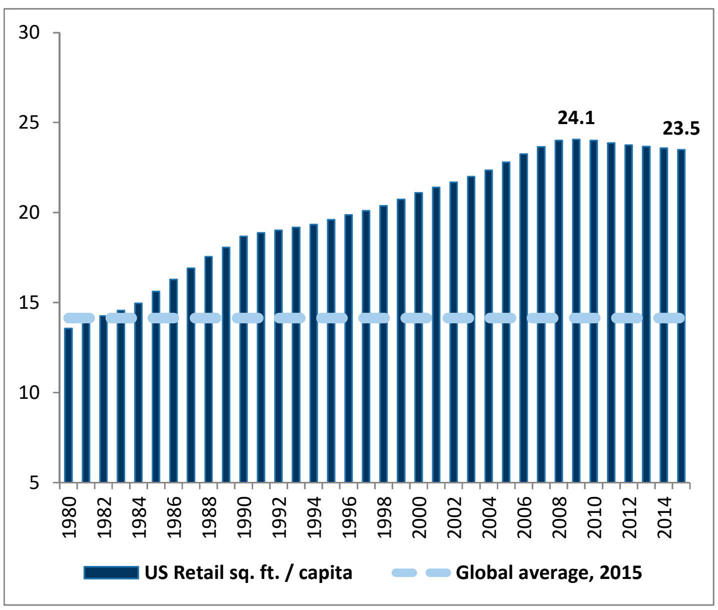
What are the pain points?

Predicting demand, trends. One of the biggest challenges within retail is navigating trends and gauging the level of demand appropriately. Specifically within apparel, designers and buyers are typically making decisions around what will be fashionable and in-demand nearly two years in advance of items hitting shelves. Current forecasting models are limited and fall short in areas such as automation, interpreting demand drivers, and limitations of historical data.

Inventory management. Inventory management remains an issue as the level of sophistication and coordination across systems often varies among members in the value chain. The effects are costly as overstocks and out-of-stocks can have a significant impact on retail sales. In the year ended spring 2015, more than $630bn in lost retail sales was attributed to out-of-stocks and more than $470bn as a result of overstocks (sales that occur at a price point where the retailer takes a loss), according to a study from IHL Group.

Number, size of stores. Store footprints, whether the amount in aggregate or per capita, remain a point of friction for retailers. In 2015, retail space reached 7.6bn sq. ft. in aggregate in the US, or 23.5 sq. ft. per capita, versus 6.7bn and 22.8 in 2005, respectively (Exhibit 55). As ecommerce continues to penetrate traditional categories such as electronics and apparel, newer categories like CPG present meaningful opportunity for greater share shift to further exacerbate the impact of surplus retail space.

Exhibit 48: US Retail space per capita
Retail square feet per capita



Source: ICSC, Euromonitor
Note: Average comprises of Australia, Brazil, Canada, China, France, Germany, India, Japan, Mexico, Russia, and the UK

What is the current way of doing business?

The current way of doing business is characterized by an extensive value chain that can be separated into 4 general buckets: production, storage, distribution, and retail. While these four steps provide a general sense of the process, within each of these buckets an additional step, partner, or intermediary can usually be found. The result is an amalgamation of systems coordinating from manufacturing through sale that can lead to overstocking, out-of-stocks, and inefficient allocation of resources – particularly during peak seasons. That said, logistics and inventory management processes have improved significantly in recent decades as more technology and systems like just-in-time manufacturing have been adopted. Third party logistics providers like UPS have also employed advanced analytics for route optimization and package management – yet another area where we see potential for AI in the future. However, the current way of doing business still presents challenges, particularly in categories like fashion, apparel, and footwear, in forecasting what consumers will want, how many, and at what price.

How does AI/ML help?

Recommendation engines. AI/ML has the potential to deepen recommendation engine capabilities by leveraging immense data sets across sales, customers, and content. One of the first opportunities in the early days of ecommerce was the recommendation engine, though most functioned based largely on product attributes as little was known about customers' preferences. Techniques like collaborative filtering have helped by leveraging known similarities in customers' preferences and tastes to make recommendations or predictions of unknown preferences.

Still, limitations such as data sparsity, or the new user/item "cold start" problem, and scalability remain issues as users grow rapidly and the consumption levels of computational resources becomes impractical. Companies like Zalando and StitchFix are already working to integrate both sales and content data with consumer preferences via machine learning, as Zalando believes customer-item affinity will ultimately drive the probability of a sale.

Customer support. Natural language processing (NLP) and image recognition also present opportunities within retail to improve customer support and stretch the parameters of traditional search. Recent acquisitions such as Etsy acquiring Blackbird Technologies, a company using AI for image recognition and NLP to deliver greater search performance, show ecommerce companies are seeking ways to improve the relevance of results and take greater advantage of the scale provided by their platforms.

NLP also presents opportunity for companies to deliver conversational user experiences and commerce. Recent Alphabet acquisition API.ai and companies like Angel.ai are creating artificially intelligent systems around natural language processing to aid commerce and customer support through both voice and messaging. In short, technologies like NLP and image recognition deliver more relevant results and service by simulating human understanding and leveraging product attributes (e.g., visual) that have historically been unobtainable.

Demand prediction and price optimization. Machine learning and AI technologies have the potential to incorporate data across customer touchpoints and content attributes to more accurately predict demand for new items and styles. Predicting demand in categories like apparel where trends can enter and exit quickly has been a long-standing challenge in retail, particularly given the runway for design and production in some categories. By leveraging AI/ML, retailers can recognize patterns and better understand local implications for promotion and price elasticity and incorporate learnings into both marketing and production processes.

Companies such as Amazon have taken steps in this direction, with the receiving of an "anticipatory package shipping" patent late in 2013. While the original filing makes no mention of machine learning, it is apparent this type of system could eventually be orchestrated by deep learning to consider not only seasonal demand but also weather, demographics, and unique user shopping patterns.

Exhibit 49: AI/ML forecasting head-to-head with traditional methods
Key strengths: Single algorithm, leverage multi-item history, optimize across processes

Traditional retail forecasts	ML/AI-based forecasts
- Handful of attributes	- Thousands of attributes
- Limited data	- Unlimited data
- Primarily time-series based	- Model identifies demand-drivers
- Crude forecasts	- Highly detailed forecasts
- Limited use	- Broad applicability
- Analyst manually enhances forecast	- Better local knowledge of demand
- Many single-purpose algorithms	**- Single algorithm**
- History for this SKU or this category only	**- Leverages history for all items/promos to forecast each item**
- Each algorithm has a restricted purpose	**- Useful in multiple processes (pricing, promo, assortment, supply chain, etc.)**
- Configuration analyst specifies	- Machine learns

Source: Predictix

Quantifying the opportunity

Improve demand prediction with reduced labor costs. Businesses in the US currently spend nearly $6bn annually in labor costs to "analyze past buying trends, sales records, price, and quality of merchandise to determine value and yield. Select, order, and authorize payment for merchandise according to contractual agreements" according to the Bureau of Labor Statistics. Said another way, wholesale and retail buyers are tasked with leveraging historical data, professional experience, fashion expertise to determine what shoppers will be interested in buying roughly two years in the future. While the continued penetration of ecommerce has increased the amount of available data for this task, the challenge remains converting that data into actionable insights that improve not just ad targeting but also the arguably more challenging exercise of predicting trends and leaning in accordingly. We believe this type of consideration could prove well-suited for AI/ML given the ability to combine not only quantitative but also visual data to predict demand and optimize buying decisions. **We estimate the integration of AI/ML practices globally could remove roughly $54bn of labor costs annually** across retail by 2025.

Exhibit 50: Labor costs associated with wholesale and retail buyers
$ millions, BPA = Buyer and Purchasing Agent

Wholesale and Retail Buyers	
US Median annual pay	$73,662
Number of jobs in US	112,868
US Annual cost ($mn)	**$8,314**
US retail spend ($mn)	$3,531,349
Global retail spend ($mn)	$22,786,689
US share	15%
Global BPA cost ($mn)	**$53,648**

Source: BLS, Euromonitor, Goldman Sachs Global Investment Research

Optimize pricing. A joint HBS and Rue La La study conducted to optimize daily pricing estimated a roughly 9.7% increase in revenue on average driven by the integration of machine learning processes, with an associated 90% confidence interval of [2.3%, 17.8%]. Given some of the nuances of a flash sales model and the volume of sales, we haircut the potential improvement by 200bps from the mean to 7.7% and assume 2.3%-7.7% improvement could be achieved by incorporating AI/ML to consider the multi-variable issue of optimizing pricing based on predicted demand. One of the challenges of dynamic pricing within retail, particularly apparel and footwear, is the lack of historical data for new styles, colors, etc. to leverage in order to predict demand. Applying machine learning, which is able to analyze hundreds of products and attributes simultaneously, will ultimately enable better evaluation and prioritization of insights from more expansive sets of data than traditional forecasting. As a result, **we see the opportunity for AI/ML driven price optimization to generate $41bn in annual sales lift on average across apparel and footwear ecommerce globally** by 2025.

Exhibit 51: AI/ML price optimization could generate significant lift in apparel ecommerce
$ millions

$ millions		2025
Global ecommerce		**$4,551,239**
% of retail		15%
Apparel/footwear ecommerce		**$815,262**
% of ecommerce		18%
		Sales lift
	2.3%	**$18,751**
	3.4%	$27,556
	4.5%	$36,361
	5.0%	**$40,763**
	5.5%	$45,165
	6.6%	$53,970
	7.7%	**$62,775**

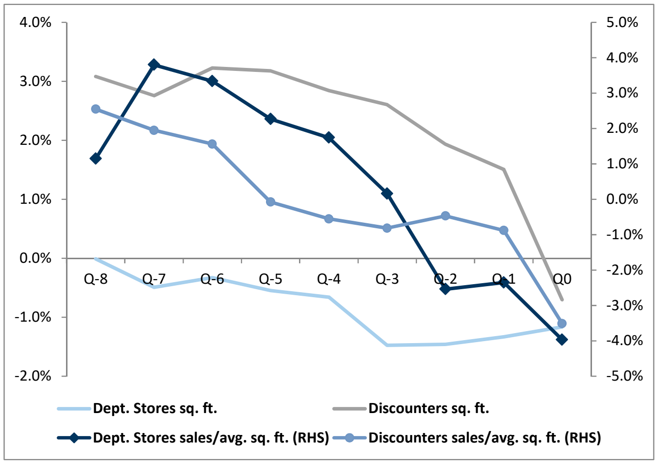
Source: Euromonitor, Goldman Sachs Global Investment Research.

Who could be disrupted?

With the integration of AI/ML across a variety of processes in the retail value chain, significant efficiencies across inventory management, production, and targeting could prove disruptive for both companies and employees. We view over-built retailers as likely to be disrupted as efficiencies within the value chain driven by AI/ML could help asset-light retailers further refine their demand prediction and inventory management at a pace ahead of larger competitors.

We also see tighter inventory management potentially proving disruptive for off-price retailers that benefit from the over-buying and/or over-production of larger retailers and brands. With more accurate production and demand forecasting, the opportunity for off-price retailers to benefit from production overruns, cancelled orders, and forecasting misjudgments could be significantly reduced. Recall, more than $470bn in sales were lost as a result of overstocks in the retail year ending spring 2015 (IHL).

Exhibit 52: Square ft. growth vs. sales/avg. sq. ft. growth
yoy % change, equal-weighted



Source: Company data

Energy

$140bn cumulative cost savings by 2025

extreme conditions. Equipment reliability is of extreme importance and failures of equipment and processes can meaningfully impact project economics. To avoid failures, the industry often over-engineers equipment and employs several layers of redundancy, raising capital employed per job or on a project. To the extent that AI/ML aids in designing more reliable equipment, industry's capex and opex requirements can be lowered. The benefits can be sizeable and **we estimate that a 1% reduction in the oil and gas industry's capex, opex and inventory management can result in savings of about $140bn over a 10-year period**.

What is the opportunity?

We believe AI/ML can help the oil and gas industry throughout the value chain:

Project planning. Major energy projects around the world can cost tens of billions of dollars and can have 3 to 5 year lead times. Managements approve these projects based on a certain set of macro assumptions, regarding oil prices and the demand and supply of their key products/services. To the extent that AI/ML applications can better inform managements about the feasibility of their projects, companies can make better decisions, reducing the number of uneconomical projects undertaken. In addition, AI/ML application can help in (1) more accurate determination of project costs, by incorporating the industry's/company's past experiences in such projects, and (2) better execution of the company's projects by keeping project costs in line with plans.

Improved equipment reliability. Unplanned equipment downtime and non-productive lost time are some of the biggest drivers of project cost escalation. The oil service industry is highly focused on improving equipment reliability, and AI/ML can help in this regard. The industry is especially targeting subsea BOPs (Blow Out Preventers) that are generally the most failure prone items on a rig and each failure can cost the deepwater industry a minimum of approx. $10mn-$15mn (see case studies below). Similarly, pressure pumping pumps are failure prone and in order to minimize lost time, service companies bring twice as many pumps to the wellside vs. what is technically required. Enhancing equipment reliability will not only reduce equipment maintenance costs but also lower a service company's capital deployed per job.

Improved identification, targeting and development of hydrocarbon resources. Finding oil and gas reserves and their exploitation generates large amounts of data. Data is generated when the industry conducts geo-seismic analysis to identify the location of oil and gas reserves. Similarly, data is generated when the well is drilled and tested. Finally, when the field is developed and produced, significant amount of production data is generated. Integrating geological data with production related data and installed hardware related data can yield information that can be used for optimal exploitation of reserves, and learnings from one project can be applied to more economical future project designs.

Increasing uptime in downstream industries. Level of planned and unplanned downtime can meaningfully impact profitability of the downstream industries. In gas pipelines, uptime on compressors is important to maintain good flow, while optimal "inspection" of pipelines can reduce unplanned downtime and leakage. Similarly, planned and unplanned downtime in the refining and petrochemical industries has a high opportunity cost. Even a 1% improvement in utilization can add up to meaningful savings.

What are the pain points?

The energy industry is highly fragmented at all levels. Nearly 400 E&Ps are involved in the exploitation of shale resources in the US, and many other upstream companies work in different parts of the world. Within the oil service industry, the Big-3 companies (Schlumberger, Halliburton and Baker Hughes) dominate most technology driven businesses, but there are many participants in the more commoditized part of services, like drilling rigs and pressure pumping. The midstream, refining and petrochemicals businesses are fragmented too.

Fragmentation creates challenges as critical data is in the hands of many players. As a result, one company may not have access to all data for a geological play or a certain type of equipment or process. Also, some companies that have access to key data may be unwilling to share it, even though they may not have the financial strength or the technical knowhow to leverage it.

Access to data. Moreover, the industry's data spans geographical boundaries, as oil and gas reserves are spread all over the world, and often data is in the hands of National Oil Companies (NOCs), which means access to data may be restricted by regulatory challenges. In addition, data spans various time periods, as the earliest well was drilled in 1880.

Finally, data analytics may be most useful when it is available across the entire value chain. But generally energy companies are primarily involved in one aspect of the business, and may not have access to all items in the value chain, that would make the analysis optimal.

Availability of data. One other pain point is availability of critical data, as in the past the industry may not have placed sensors in key parts of the hydrocarbon chain that would help AI/ML applications. As an example, while the industry may have critical data on how often a BOP breaks down and what pressure it encountered during its operational life, it may not have data on the temperature, current and voltage readings on various coils and electronic components within the BOP. The industry is now starting to put such sensors on new products, and it will take some time before companies will have this additional data.

What is the current way of doing business?

The industry is still using conventional methods for exploiting oil and gas reserves and is using methods and techniques that are evolutionary but not really revolutionary. The key issue afflicting the industry is that the industry is being run in various silos and there is limited integration and cohesiveness across various parts of the business. For example, the owners of the oil and gas reserves (E&Ps) design the whole project, and then divide the work among discrete service providers. E&Ps have the most information, but they are not well versed on what service companies can offer, and often they keep service companies at an arm's length, believing that undue reliance on them could lead to cost increases in future and leakage of their IP. For the energy industry to really gain from AI/ML, data will need to be more widely shared between E&Ps and service sector, and a more collaborative model needs to emerge. In the offshore space, as IOCs have struggled with costs, some IOCs (International Oil Companies) have taken leadership in collaborating with well integrated service companies.

How does AI/ML help?

AI/ML can help in several ways:

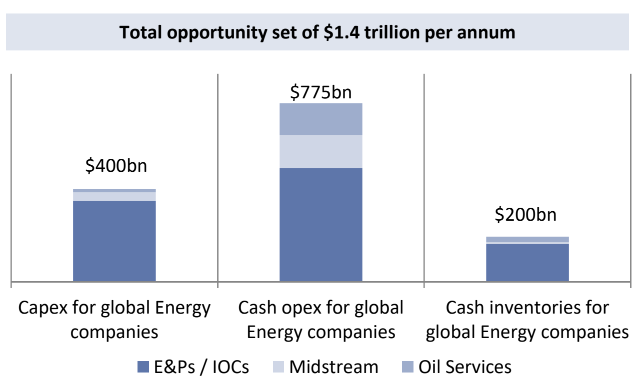
- Improving reliability of products by incorporating knowledge gained from historical information. AI/ML can also help in reducing the time between product development, field trials and commercialization.
- Better targeting of the oil and gas reserves, by cutting the time and cost to drill wells, leading to lower field development costs.
- Lowering production costs during the life of a field, through better equipment uptime and reduced maintenance costs.
- Improved uptime for offshore and land rigs, resulting in higher drilling efficiency and reduction in days to drill a well.
- AI/ML based data analysis can lower maintenance related downtime in the downstream industries.

Quantifying the opportunity

In 2016, we expect GS covered oil and gas companies to spend nearly $400bn in capex. In addition, the oil and gas industry should spend another $775 bn in operating costs (excluding cost of goods sold for refining and petrochemicals and DD&A), and hold about $200bn in inventory.

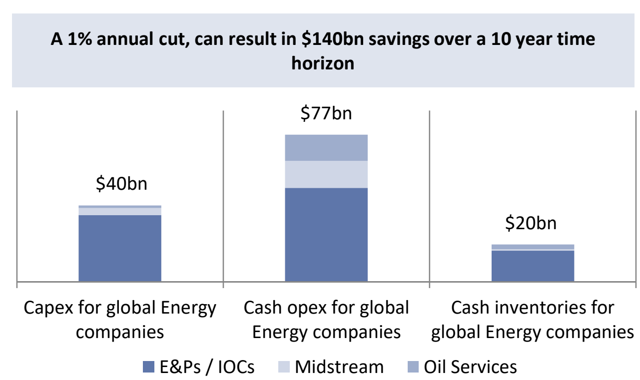
We estimate that should AI/ML applications lower capex and opex by 1%, and the industry reduces its inventory by 1% through better inventory management, total savings for the industry over a 10-year period would amount to $140bn. We present below several case studies that point to areas where costs could be reduced.

Exhibit 53: GS covered energy companies spend $1.4tn each year on capex + opex + inventories
For IOCs, opex figures are only for the upstream business and not for downstream. Data excludes NOCs and not covered public companies



Source: Goldman Sachs Global Investment Research.

Exhibit 54: We see $140bn in savings over 10-years for a 1% cut in capex, operating costs and inventory
For IOCs, opex figures are only for the upstream business and not for downstream. Data excludes NOCs and not covered public companies

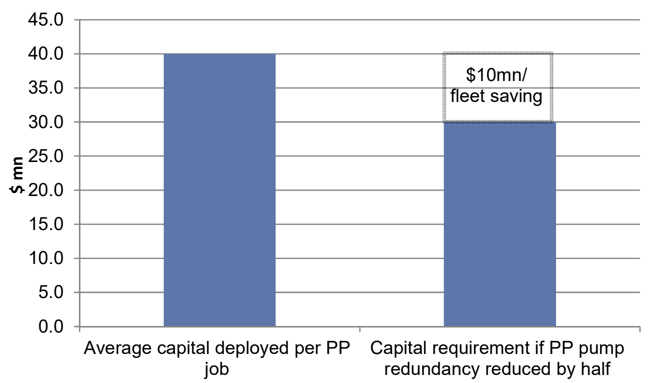


Source: Goldman Sachs Global Investment Research.

Reducing pressure pumping fleet costs. The industry's pressure pumping fleet experiences very high equipment attrition and equipment maintenance typically runs about 10%-15% of the cost of revenues. Over the past 5-years, pressure pumping on average has been a $30bn annual revenue business, and the industry has spent nearly $3.6bn annually in maintaining the pressure pumping equipment, a figure which excludes major capital costs like equipment upgrades etc.

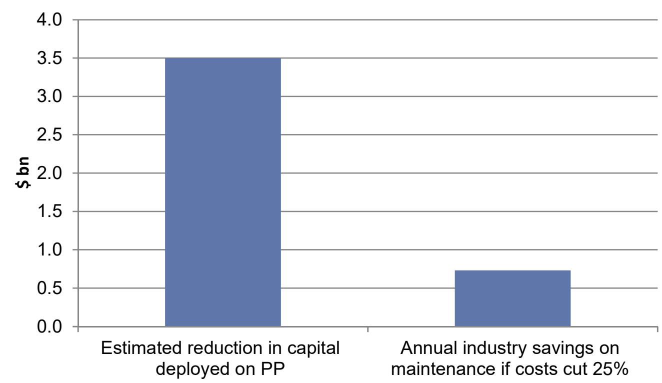
An average pressure pumping job in the US requires about 20,000 HHP. However, the industry typically takes about 40,000 HHP to the wellsite, in order to maintain redundancy and lower non-productive job time in case of equipment failure. This level of redundancy is inefficient for the oil and gas industry. Should equipment reliability be enhanced, the level of equipment redundancy needed at the wellsite will be reduced. We estimate that if the level of redundancy required in the field is reduced in half, the capital deployed on a pressure pumping job could be reduced by 25% from about $40mn to $30mn. Similarly, **predictive analytics can reduce equipment maintenance, and we estimate that a 25% reduction in maintenance costs could lead to the industry saving about $0.7bn annually (or $7bn in 10-years) on the industry's fleet of 14mn HHP at 85% equipment utilization.**

Exhibit 55: Avg. capital deployed per PP fleet could be cut by 25%
Assuming 40,000HHP average US pressure pumping fleet



Source: Goldman Sachs Global Investment Research.

Exhibit 56: The industry could cut total capital deployed in PP fleets by $3.5bn
25% lower maintenance costs is a $0.7bn annual opportunity



Source: Goldman Sachs Global Investment Research.

Improving drilling time through "Rig of the Future". The oil and gas industry spends $0.7mn to $1.0mn per day when a deepwater well is being drilled, while daily drilling costs on horizontal shale wells can be around $60K/day. The industry can thus generate sizeable savings for each day that it shaves from its drilling program.

Drilling times can be reduced on a drilling rig in three ways:

- Improving equipment uptime, especially on problem items like BOPs (especially subsea) and Top Drives.
- Choosing the right bottom-hole-assembly for conditions likely to be encountered in a well.
- Optimizing drilling performance, by establishing a "closed loop information system" between surface equipment and the "bottom hole assembly". Automating systems and reducing the impact of "crew quality" in generating repeatable good performance.

The industry is actively working on reducing downtime on BOPs (Blow Out Preventers) and Top Drives, by sifting through data, and looking for leading edge signals that predict upcoming problems.

Similarly, data analysis from previously drilled wells can help oil companies design optimal drilling fluids and drilling bits for a particular well. In addition, by establishing a closed loop system between sensors near the drill bit, and the controls on a driller's panel, a "Smart Rig" can be designed, which automatically adjusts "weight-on-bit" and the torque applied on the drill string, to most efficiently drill a well based on encountered downhole conditions. AI/ML can help in continuous improvement and repeatable performance.

One of the key issues in drilling is the impact of "human intervention", and the industry has found that even on similar wells, drilling performance can vary significantly depending on the quality of the crew. Automation can reduce the impact of "human intervention" on drilling performance.

Improving refinery uptime. The US refining industry has an installed base of roughly 18mn bpd, and enjoys an average utilization of about 90%, as 10% of the time the industry undergoes planned and unplanned maintenance. On average, refining margins are about $10/b, meaning the industry forgoes $10/b for each barrel of production that is offline. We estimate that if through better data analytics the industry can reduce maintenance related downtime from 10% to 9%, refiners would realize an additional margin of $657mn annually or about $6.6bn in ten years.

Exhibit 57: Reduction in maintenance related downtime in refineries from 10% to 9% can save US refiners $6.6bn over 10 years
US refiners on average give up $6.6bn in margins each year owing to 10% average downtime

Global refining capacity (mn bpd)	18.0
Avg. refinery downtime during a year	10%
Implied annual refinery capacity down (mn bpd)	1.8
Avg. refinery margin per bbl processed	$10.0
Lost revenue per year due to avg. 10% refinery downtime ($ bn)	6.6
Lost revenue per year @ avg. 9% downtime ($ bn)	5.9
Annual savings from 100bp improvement in downtime ($ bn)	**0.7**
Total savings over a 10 year horizon ($ bn)	**6.6**

Source: Goldman Sachs Global Investment Research.

Who could be disrupted?

Small / less sophisticated energy companies or companies that are capital constrained and have limited technical know-how would be the most affected negatively, as better positioned companies employ AI/ML to lower costs. This will be equally true for the E&P and the oil service sector. Key winners will be those who invested in the past in acquiring data from their assets and had the foresight to store it. These will also be companies that not only have the financial ability to employ AI/ML techniques to leverage this data, but also have a culture of technology usage and innovation to leverage these new data analytics techniques.

Disclosure Appendix

Reg AC

We, Heath P. Terry, CFA, Jesse Hulsing, Piyush Mubayi and Waqar Syed, hereby certify that all of the views expressed in this report accurately reflect our personal views about the subject company or companies and its or their securities. We also certify that no part of our compensation was, is or will be, directly or indirectly, related to the specific recommendations or views expressed in this report.

Unless otherwise stated, the individuals listed on the cover page of this report are analysts in Goldman Sachs' Global Investment Research division.

Investment Profile

The Goldman Sachs Investment Profile provides investment context for a security by comparing key attributes of that security to its peer group and market. The four key attributes depicted are: growth, returns, multiple and volatility. Growth, returns and multiple are indexed based on composites of several methodologies to determine the stocks percentile ranking within the region's coverage universe.

The precise calculation of each metric may vary depending on the fiscal year, industry and region but the standard approach is as follows:

Growth is a composite of next year's estimate over current year's estimate, e.g. EPS, EBITDA, Revenue. **Return** is a year one prospective aggregate of various return on capital measures, e.g. CROCI, ROACE, and ROE. **Multiple** is a composite of one-year forward valuation ratios, e.g. P/E, dividend yield, EV/FCF, EV/EBITDA, EV/DACF, Price/Book. **Volatility** is measured as trailing twelve-month volatility adjusted for dividends.

Quantum

Quantum is Goldman Sachs' proprietary database providing access to detailed financial statement histories, forecasts and ratios. It can be used for in-depth analysis of a single company, or to make comparisons between companies in different sectors and markets.

GS SUSTAIN

GS SUSTAIN is a global investment strategy aimed at long-term, long-only performance with a low turnover of ideas. The GS SUSTAIN focus list includes leaders our analysis shows to be well positioned to deliver long term outperformance through sustained competitive advantage and superior returns on capital relative to their global industry peers. Leaders are identified based on quantifiable analysis of three aspects of corporate performance: cash return on cash invested, industry positioning and management quality (the effectiveness of companies' management of the environmental, social and governance issues facing their industry).

Disclosures

Coverage group(s) of stocks by primary analyst(s)

Heath P. Terry, CFA: America-Internet. Jesse Hulsing: America-Emerging Software. Piyush Mubayi: Asia Pacific Telecoms, China Internet. Waqar Syed: America-Offshore Drilling Services, America-Oilfield Services & Equipment.

America-Emerging Software: Apptio Inc., Benefitfocus Inc., Cornerstone OnDemand Inc., Guidewire Software Inc., Hortonworks Inc., HubSpot Inc., Intuit Inc., New Relic Inc., ServiceNow Inc., Shopify Inc., Splunk Inc., Tableau Software, Talend SA, Teradata Corp., Ultimate Software Group, Veeva Systems Inc., Zendesk Inc..

America-Internet: Amazon.com Inc., Bankrate Inc., Criteo SA, eBay Inc., Endurance International Group, Etsy Inc., Expedia Inc., Groupon Inc., GrubHub Inc., IAC/InterActiveCorp, LendingClub Corp., LinkedIn Corp., Match Group, Netflix Inc., Pandora Media Inc., PayPal Holdings, Priceline.com Inc., Shutterfly Inc., TripAdvisor Inc., TrueCar, Twitter Inc., WebMD Health Corp., Yahoo! Inc., Yelp Inc., Zillow Group, Zynga Inc..

America-Offshore Drilling Services: Atwood Oceanics Inc., Diamond Offshore Drilling, Ensco Plc, Noble Corp., Oceaneering International Inc., Rowan Cos., Transocean Ltd..

America-Oilfield Services & Equipment: Baker Hughes Inc., Carbo Ceramics Inc., Emerge Energy Services LP, Fairmount Santrol Holdings, FMC Technologies, Frank's International NV, Halliburton Co., Helmerich & Payne Inc., Nabors Industries Ltd., National Oilwell Varco, Patterson-UTI Energy Inc., Pioneer Energy Services Corp., RPC Inc., Schlumberger Ltd., Weatherford International Ltd..

Asia Pacific Telecoms: Advanced Info Service PCL, Astro Malaysia Holdings, Axiata Group, Chunghwa Telecom, Digi.com, Far EasTone, HKT Trust, Hong Kong Broadband Network Ltd., Indosat, Intouch Holdings, KT Corp., KT Corp. (ADR), LG UPlus, M1 Ltd., Maxis Bhd, PCCW Ltd., PT Link Net Tbk, PT Sarana Menara Nusantara, PT XL Axiata, Singapore Telecommunications, SK Telecom, SK Telecom (ADR), SmarTone, StarHub, Taiwan Mobile, Telekom Malaysia, Telekomunikasi Indonesia, Total Access Communications, Tower Bersama Infrastructure Tbk, True Corp.

China Internet: 58.com Inc., Alibaba Group, Baidu.com Inc., Ctrip.com International, Gridsum, JD.com Inc., NetEase Inc., New Oriental Education & Technology, SINA Corp., TAL Education Group, Tencent Holdings, Vipshop Holdings, Weibo Corp..

Distribution of ratings/investment banking relationships

Goldman Sachs Investment Research global Equity coverage universe

	Rating Distribution			Investment Banking Relationships		
	Buy	Hold	Sell	Buy	Hold	Sell
Global	31%	55%	14%	64%	59%	53%

As of October 1, 2016, Goldman Sachs Global Investment Research had investment ratings on 2,921 equity securities. Goldman Sachs assigns stocks as Buys and Sells on various regional Investment Lists; stocks not so assigned are deemed Neutral. Such assignments equate to Buy, Hold and Sell for the purposes of the above disclosure required by the FINRA Rules. See 'Ratings, Coverage groups and views and related definitions' below. The Investment Banking Relationships chart reflects the percentage of subject companies within each rating category for whom Goldman Sachs has provided investment banking services within the previous twelve months.

Regulatory disclosures

Disclosures required by United States laws and regulations

See company-specific regulatory disclosures above for any of the following disclosures required as to companies referred to in this report: manager or co-manager in a pending transaction; 1% or other ownership; compensation for certain services; types of client relationships; managed/co-managed public offerings in prior periods; directorships; for equity securities, market making and/or specialist role. Goldman Sachs trades or may trade as a principal in debt securities (or in related derivatives) of issuers discussed in this report.

The following are additional required disclosures: **Ownership and material conflicts of interest:** Goldman Sachs policy prohibits its analysts, professionals reporting to analysts and members of their households from owning securities of any company in the analyst's area of coverage. **Analyst compensation:** Analysts are paid in part based on the profitability of Goldman Sachs, which includes investment banking revenues. **Analyst as officer or director:** Goldman Sachs policy prohibits its analysts, persons reporting to analysts or members of their households from serving as an officer, director, advisory board member or employee of any company in the analyst's area of coverage. **Non-U.S. Analysts:** Non-U.S. analysts may not be associated persons of Goldman, Sachs & Co. and therefore may not be subject to FINRA Rule 2241 or FINRA Rule 2242 restrictions on communications with subject company, public appearances and trading securities held by the analysts.

Distribution of ratings: See the distribution of ratings disclosure above. **Price chart:** See the price chart, with changes of ratings and price targets in prior periods, above, or, if electronic format or if with respect to multiple companies which are the subject of this report, on the Goldman Sachs website at http://www.gs.com/research/hedge.html.

Additional disclosures required under the laws and regulations of jurisdictions other than the United States

The following disclosures are those required by the jurisdiction indicated, except to the extent already made above pursuant to United States laws and regulations. **Australia:** Goldman Sachs Australia Pty Ltd and its affiliates are not authorised deposit-taking institutions (as that term is defined in the Banking Act 1959 (Cth)) in Australia and do not provide banking services, nor carry on a banking business, in Australia. This research, and any access to it, is intended only for "wholesale clients" within the meaning of the Australian Corporations Act, unless otherwise agreed by Goldman Sachs. In producing research reports, members of the Global Investment Research Division of Goldman Sachs Australia may attend site visits and other meetings hosted by the issuers the subject of its research reports. In some instances the costs of such site visits or meetings may be met in part or in whole by the issuers concerned if Goldman Sachs Australia considers it is appropriate and reasonable in the specific circumstances relating to the site visit or meeting. **Brazil:** Disclosure information in relation to CVM Instruction 483 is available at http://www.gs.com/worldwide/brazil/area/gir/index.html. Where applicable, the Brazil-registered analyst primarily responsible for the content of this research report, as defined in Article 16 of CVM Instruction 483, is the first author named at the beginning of this report, unless indicated otherwise at the end of the text. **Canada:** Goldman Sachs Canada Inc. is an affiliate of The Goldman Sachs Group Inc. and therefore is included in the company specific disclosures relating to Goldman Sachs (as defined above). Goldman Sachs Canada Inc. has approved of, and agreed to take responsibility for, this research report in Canada if and to the extent that Goldman Sachs Canada Inc. disseminates this research report to its clients. **Hong Kong:** Further information on the securities of covered companies referred to in this research may be obtained on request from Goldman Sachs (Asia) L.L.C. **India:** Further information on the subject company or companies referred to in this research may be obtained from Goldman Sachs (India) Securities Private Limited, Research Analyst - SEBI Registration Number INH000001493, 951-A, Rational House, Appasaheb Marathe Marg, Prabhadevi, Mumbai 400 025, India, Corporate Identity Number U74140MH2006FTC160634, Phone +91 22 6616 9000, Fax +91 22 6616 9001. Goldman Sachs may beneficially own 1% or more of the securities (as such term is defined in clause 2 (h) the Indian Securities Contracts (Regulation) Act, 1956) of the subject company or companies referred to in this research report. **Japan:** See below. **Korea:** Further information on the subject company or companies referred to in this research may be obtained from Goldman Sachs (Asia) L.L.C., Seoul Branch. **New Zealand:** Goldman Sachs New Zealand Limited and its affiliates are neither "registered banks" nor "deposit takers" (as defined in the Reserve Bank of New Zealand Act 1989) in New Zealand. This research, and any access to it, is intended for "wholesale clients" (as defined in the Financial Advisers Act 2008) unless otherwise agreed by Goldman Sachs. **Russia:** Research reports distributed in the Russian Federation are not advertising as defined in the Russian legislation, but are information and analysis not having product promotion as their main purpose and do not provide appraisal within the meaning of the Russian legislation on appraisal activity. **Singapore:** Further information on the covered companies referred to in this research may be obtained from Goldman Sachs (Singapore) Pte. (Company Number: 198602165W). **Taiwan:** This material is for reference only and must not be reprinted without permission. Investors should carefully consider their own investment risk. Investment results are the responsibility of the individual investor. **United Kingdom:** Persons who would be categorized as retail clients in the United Kingdom, as such term is defined in the rules of the Financial Conduct Authority, should read this research in conjunction with prior Goldman Sachs research on the covered companies referred to herein and should refer to the risk warnings that have been sent to them by Goldman Sachs International. A copy of these risks warnings, and a glossary of certain financial terms used in this report, are available from Goldman Sachs International on request.

European Union: Disclosure information in relation to Article 4 (1) (d) and Article 6 (2) of the European Commission Directive 2003/125/EC is available at http://www.gs.com/disclosures/europeanpolicy.html which states the European Policy for Managing Conflicts of Interest in Connection with Investment Research.

Japan: Goldman Sachs Japan Co., Ltd. is a Financial Instrument Dealer registered with the Kanto Financial Bureau under registration number Kinsho 69, and a member of Japan Securities Dealers Association, Financial Futures Association of Japan and Type II Financial Instruments Firms Association. Sales and purchase of equities are subject to commission pre-determined with clients plus consumption tax. See company-specific disclosures as to any applicable disclosures required by Japanese stock exchanges, the Japanese Securities Dealers Association or the Japanese Securities Finance Company.

Ratings, coverage groups and views and related definitions

Buy (B), Neutral (N), Sell (S) -Analysts recommend stocks as Buys or Sells for inclusion on various regional Investment Lists. Being assigned a Buy or Sell on an Investment List is determined by a stock's return potential relative to its coverage group as described below. Any stock not assigned as a Buy or a Sell on an Investment List is deemed Neutral. Each regional Investment Review Committee manages various regional Investment Lists to a global guideline of 25%-35% of stocks as Buy and 10%-15% of stocks as Sell; however, the distribution of Buys and Sells in any particular coverage group may vary as determined by the regional Investment Review Committee. Regional Conviction Buy and Sell lists represent investment recommendations focused on either the size of the potential return or the likelihood of the realization of the return.

Return potential represents the price differential between the current share price and the price target expected during the time horizon associated with the price target. Price targets are required for all covered stocks. The return potential, price target and associated time horizon are stated in each report adding or reiterating an Investment List membership.

Coverage groups and views: A list of all stocks in each coverage group is available by primary analyst, stock and coverage group at http://www.gs.com/research/hedge.html. The analyst assigns one of the following coverage views which represents the analyst's investment outlook on the coverage group relative to the group's historical fundamentals and/or valuation. **Attractive (A).** The investment outlook over the following 12 months is favorable relative to the coverage group's historical fundamentals and/or valuation. **Neutral (N).** The investment outlook over the following 12 months is neutral relative to the coverage group's historical fundamentals and/or valuation. **Cautious (C).** The investment outlook over the following 12 months is unfavorable relative to the coverage group's historical fundamentals and/or valuation.

Not Rated (NR). The investment rating and target price have been removed pursuant to Goldman Sachs policy when Goldman Sachs is acting in an advisory capacity in a merger or strategic transaction involving this company and in certain other circumstances. **Rating Suspended (RS).** Goldman Sachs Research has suspended the investment rating and price target for this stock, because there is not a sufficient fundamental basis for determining, or there are legal, regulatory or policy constraints around publishing, an investment rating or target. The previous investment rating and price target, if any, are no longer in effect for this stock and should not be relied upon. **Coverage Suspended (CS).** Goldman Sachs has suspended coverage of this company. **Not Covered (NC).** Goldman Sachs does not cover this company. **Not Available or Not Applicable (NA).** The information is not available for display or is not applicable. **Not Meaningful (NM).** The information is not meaningful and is therefore excluded.

Global product; distributing entities

The Global Investment Research Division of Goldman Sachs produces and distributes research products for clients of Goldman Sachs on a global basis. Analysts based in Goldman Sachs offices around the world produce equity research on industries and companies, and research on macroeconomics, currencies, commodities and portfolio strategy. This research is disseminated in Australia by Goldman Sachs Australia Pty Ltd (ABN 21 006 797 897); in Brazil by Goldman Sachs do Brasil Corretora de Títulos e Valores Mobiliários S.A.; in Canada by either Goldman Sachs Canada Inc. or Goldman, Sachs & Co.; in Hong Kong by Goldman Sachs (Asia) L.L.C.; in India by Goldman Sachs (India) Securities Private Ltd.; in Japan by Goldman Sachs Japan Co., Ltd.; in the Republic of Korea by Goldman Sachs (Asia) L.L.C., Seoul Branch; in New Zealand by Goldman Sachs New Zealand Limited; in Russia by OOO Goldman Sachs; in Singapore by Goldman Sachs (Singapore) Pte. (Company Number: 198602165W); and in the United States of America by Goldman, Sachs & Co. Goldman Sachs International has approved this research in connection with its distribution in the United Kingdom and European Union.

European Union: Goldman Sachs International authorised by the Prudential Regulation Authority and regulated by the Financial Conduct Authority and the Prudential Regulation Authority, has approved this research in connection with its distribution in the European Union and United Kingdom; Goldman Sachs AG and Goldman Sachs International Zweigniederlassung Frankfurt, regulated by the Bundesanstalt für Finanzdienstleistungsaufsicht, may also distribute research in Germany.

General disclosures

This research is for our clients only. Other than disclosures relating to Goldman Sachs, this research is based on current public information that we consider reliable, but we do not represent it is accurate or complete, and it should not be relied on as such. The information, opinions, estimates and forecasts contained herein are as of the date hereof and are subject to change without prior notification. We seek to update our research as appropriate, but various regulations may prevent us from doing so. Other than certain industry reports published on a periodic basis, the large majority of reports are published at irregular intervals as appropriate in the analyst's judgment.

Goldman Sachs conducts a global full-service, integrated investment banking, investment management, and brokerage business. We have investment banking and other business relationships with a substantial percentage of the companies covered by our Global Investment Research Division. Goldman, Sachs & Co., the United States broker dealer, is a member of SIPC (http://www.sipc.org).

Our salespeople, traders, and other professionals may provide oral or written market commentary or trading strategies to our clients and principal trading desks that reflect opinions that are contrary to the opinions expressed in this research. Our asset management area, principal trading desks and investing businesses may make investment decisions that are inconsistent with the recommendations or views expressed in this research.

The analysts named in this report may have from time to time discussed with our clients, including Goldman Sachs salespersons and traders, or may discuss in this report, trading strategies that reference catalysts or events that may have a near-term impact on the market price of the equity securities discussed in this report, which impact may be directionally counter to the analyst's published price target expectations for such stocks. Any such trading strategies are distinct from and do not affect the analyst's fundamental equity rating for such stocks, which rating reflects a stock's return potential relative to its coverage group as described herein.

We and our affiliates, officers, directors, and employees, excluding equity and credit analysts, will from time to time have long or short positions in, act as principal in, and buy or sell, the securities or derivatives, if any, referred to in this research.

The views attributed to third party presenters at Goldman Sachs arranged conferences, including individuals from other parts of Goldman Sachs, do not necessarily reflect those of Global Investment Research and are not an official view of Goldman Sachs.

Any third party referenced herein, including any salespeople, traders and other professionals or members of their household, may have positions in the products mentioned that are inconsistent with the views expressed by analysts named in this report.

This research is not an offer to sell or the solicitation of an offer to buy any security in any jurisdiction where such an offer or solicitation would be illegal. It does not constitute a personal recommendation or take into account the particular investment objectives, financial situations, or needs of individual clients. Clients should consider whether any advice or recommendation in this research is suitable for their particular circumstances and, if appropriate, seek professional advice, including tax advice. The price and value of investments referred to in this research and the income from them may fluctuate. Past performance is not a guide to future performance, future returns are not guaranteed, and a loss of original capital may occur. Fluctuations in exchange rates could have adverse effects on the value or price of, or income derived from, certain investments.

Certain transactions, including those involving futures, options, and other derivatives, give rise to substantial risk and are not suitable for all investors. Investors should review current options disclosure documents which are available from Goldman Sachs sales representatives or at http://www.theocc.com/about/publications/character-risks.jsp. Transaction costs may be significant in option strategies calling for multiple purchase and sales of options such as spreads. Supporting documentation will be supplied upon request.

All research reports are disseminated and available to all clients simultaneously through electronic publication to our internal client websites. Not all research content is redistributed to our clients or available to third-party aggregators, nor is Goldman Sachs responsible for the redistribution of our research by third party aggregators. For research, models or other data available on a particular security, please contact your sales representative or go to http://360.gs.com.

Disclosure information is also available at http://www.gs.com/research/hedge.html or from Research Compliance, 200 West Street, New York, NY 10282.